

A timely boost to research on carbon sequestration

Removing carbon dioxide from the atmosphere is as important as controlling emissions. Basic and applied research need to be pursued in the face of political obstacles.

Alongside attempts to limit mankind's production of greenhouse gases, there is a pressing need to find ways of removing carbon from the atmosphere and 'sequestering' it in the land, in geological formations and in the oceans. The US government has now taken laudable steps in that direction, but there is a danger that progress could continue to be unjustifiably hindered.

Nearly two years ago, the US President's Committee of Advisors on Science and Technology issued a report that included a recommendation for the US Department of Energy (DoE) to increase research and development in carbon sequestration. The DoE has now announced the creation of two centres to examine ocean and terrestrial sequestration methods. It is also soliciting proposals for research, to the tune of at least \$18 million, towards the separation and capture of carbon dioxide from the atmosphere, as well as sequestration (see page 315).

Hitherto, the political climate in Washington has kept this research pinned to the drawing-board. Some politicians in Congress oppose anything suggesting that the United States might be conforming with the Kyoto Protocol negotiated to address global warming. Politicians also see the projected costs of sequestration, and fear the drag on the US economy. Meanwhile, environmentalists have been slow to embrace sequestration research, believing it might undermine hard-won gains in renewable energy and conservation efforts.

At a DoE workshop on sequestration last week, researchers were bullish about ocean and geological sequestration research. More positively still, there is an encouraging level of cooperation between the agency's Office of Science and Office of Fossil Energy — offices that haven't always had a smooth working relationship. But on sequestration, the Office of Science's orientation towards more basic research appears to fit well alongside the more applied outlook of the Office of

Fossil Energy. Such cooperation will be necessary to lead the way as the complexities and costs of sequestration become better understood. This cooperative attitude must also extend to the partnerships between scientists and industry if large-scale sequestration is to be made reality.

For significant sequestration to be feasible, much more basic research must be conducted, particularly in the oceans. Already, some companies are fertilizing the ocean with nutrients to increase fish harvests while trumpeting the additional benefits of sequestration. Research on ocean 'scrubbing' (enhanced rates of removal by organisms) as well as carbon dioxide 'injection' (disposal in liquid form for hundreds of years at depths below one kilometre or so) needs to come before the application of such commercial efforts, especially given the possible ecological consequences of such practices.

That is why a planned experiment off Hawaii is so important, and requires active support. Scientists at the Pacific International Center for High Technology Research, collaborating with an international team, are planning to inject liquefied carbon dioxide at a depth of about 900 metres off the Kona coast. The experiment, proposed for the Natural Energy Laboratory of Hawaii, would be conducted in 2001 in an established research corridor in the ocean. It is designed to gather the environmental data to properly assess future sequestration experiments. The authority that governs the Natural Energy Laboratory has yet to approve the experiment. But some environmental groups fear that it will harm the ocean ecosystem and give encouragement to the construction of another power plant. Scientists fear that the entire experiment may be jeopardized by delays. If the opportunity to conduct a reasoned environmental-assessment experiment disappears, both environment and science will be the losers. ■

Reclusion to be discouraged

A survey has revealed that many Japanese researchers are disconnected from the outside world. Their institutions can help.

In 1998, *Nature* published a total of 84 papers from Japan — the highest number ever. That record is likely to be exceeded this year. This continually upward trend would seem to indicate that Japanese researchers are taking a more active role in the international science arena. However, a report released last week by the Science and Technology Agency states otherwise (see page 314).

The report, based on a survey across a wide range of research institutions in Japan, reveals that a significant proportion of Japanese researchers do not publish internationally. According to the report, 40 per cent of researchers submit papers only to Japanese journals. The reason behind this, say researchers who responded to the survey, is the fact that many institutions still place a strong emphasis on the numerical output of research papers as a criterion for evaluating research performance. Such a system forces many Japanese researchers, including highly talented ones, to shy away from international journals with high impact factors and to turn instead to

Japanese publications, which provide a better opportunity for their papers to get published.

But the system alone cannot be blamed for Japanese scientists' lack of involvement in international science. They themselves need to make an effort to enhance their communication with researchers in other countries. The survey reveals that only 20 per cent of researchers communicate regularly with overseas colleagues. A large percentage of them do not communicate at all, and less than half use e-mail and faxes for their communication.

In the age of the Internet, there is every reason to encourage more proactivity on the part of both Japanese and Western scientists in establishing contacts, despite the obstacles posed by language. And for Japanese scientists to rise more frequently to the challenges of international scientific communication, their institutions need to focus more on quality rather than quantity in their assessment of scientific output. ■





China

Innovation is the key to prosperity, say China's leaders
p312



Space

Japan's Subaru telescope puts Ring Nebula in focus
p314



Global warming

How CO₂ uptake by ecosystems could cut global warming
p315



Museums

Research at German science museums threatened by cuts
p315

Gene estimate rises as US and UK discuss freedom of access

London

One of the leading private-sector participants in US genome sequencing efforts says he has firm evidence that there may be more than 140,000 genes in the human genome — a significant increase over conventional estimates closer to 100,000.

Randall Scott, president and chief scientific officer of the biotechnology company Incyte Pharmaceuticals, suggested this new figure during a presentation on Monday (20 September) to the annual sequencing conference organized in Miami, Florida, by the Institute for Genomic Research (TIGR).

His announcement coincides with the news that the chief scientific advisers to the US and British governments, Neal Lane and Robert May respectively, have been discussing a joint declaration underlining the commitment of their governments to public access to raw sequence data.

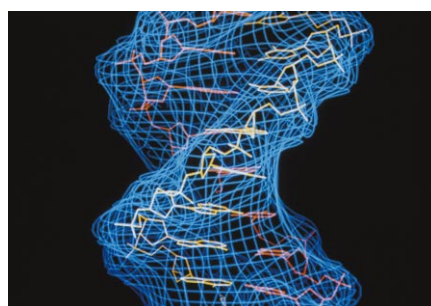
Scott's new estimate of the number of human genes will come as little surprise to most researchers. Already one result of sequencing work on other organisms, for example the fruitfly *Drosophila*, has been to increase the estimate of the number of genes these organisms contain, usually by a factor of about 20 per cent.

Nevertheless, the higher number will be of considerable interest to geneticists, particularly as it follows other suggestions that the total number of nucleotide bases in the genome is considerably more than the figure of three billion usually quoted in debates on sequencing projects.

Scott's new estimate of the total number of human genes is based on an analysis of the prevalence in genes of CpG islands — short stretches of DNA that can be methylated and as such provide a mechanism for controlling gene expression (for full details see <http://www.incyte.com>).

Incyte has already produced a large bank of sequence data that is made available for searching by other companies and research institutions on a contract basis. Incyte researchers have sequenced just under half of the CpG islands and compared them to the finished sequences of genes now available.

"This has allowed us to estimate that, over-



How many genes in the human genome? 142,634 according to Incyte Pharmaceuticals.

all, 53 per cent of genes have CpG islands associated with them," says Scott. A further estimation that there are just over 75,000 CpG islands in total in the genome has led Scott and his team to predict a total of 142,634 genes.

"Previous recent estimates appear to have been substantially lower than this because they have overestimated the frequency of CpG islands in genes," he says. "One of the implications is that the genome is even more complex than we originally thought."

Incyte's calculations are likely to increase interest in the question of patenting of sequencing data, which produced headlines

in Britain this week when *The Guardian* revealed the discussions between Washington and London on a possible joint declaration on access to sequence data. The report was later confirmed by the Department of Trade and Industry.

Officials on both sides of the Atlantic were quick to point out, however, that such a declaration would be aimed primarily at ensuring rapid access to raw sequence data, rather than preventing the patenting of genetic data as such (including the right to patent a specific gene when its sequence data is linked to a specified application).

The move has been welcomed by Britain's Wellcome Trust, which is sponsoring one third of the total human genome sequencing effort. The trust has been insistent — together with its US partners — that one condition of its support is that all such data is made publicly available within 24 hours.

John Sulston, director of the Sanger Centre near Cambridge, which is jointly funded by Wellcome and the Medical Research Council, says: "This data must be shared and controlled by all, and I strongly endorse the idea of giving the human genome an 'international ownership' flavour in this way." **David Dickson**

Putting a price on research reading

Paris

Journals produced by not-for-profit organizations are generally better value for money than those of commercial publishers. This is the conclusion of a study by the University of Wisconsin library of the costs and impacts of journals in physics, neuroscience and economics.

The data were collected last year to commemorate a famous study carried out a decade earlier by Henry Barschall for the American Institute of Physics. The original study came to prominence when the Swiss-based publisher Gordon & Breach, whose journals fared badly in the study, sued the institute (whose journals came off well).

Earlier this year, the US Court of Appeals for the Second Circuit threw out Gordon &

Breach's case, concluding that the study was reliable, and that publication and promotional use of a survey of journal prices did not constitute false or misleading advertising (see *Nature* 397, 465; 1999). Barschall died before the verdict was known.

The new study confirms that the trends highlighted by Barschall still exist, for example, the wide variation in the cost of journals per 1,000 characters. Physics journals varied by a factor of 36, for example, from 0.76 cents to 27.33 cents. Even greater variation was found when journals were assessed on their 'impact factor', a value ranging from 0.2 to 182 cents.

Of the three fields included in the study, the average cost per 1,000 characters was lowest in physics at 9.84

cents; the cost for economics journals was 10.60 and that for neuroscience journals 13.83.

But in terms of cost-effectiveness, as measured by the ratio of cost to impact, neuroscience, at 7.7, emerged better than physics journals (11.5) and much better than economics publications (29.8); a lower ratio denotes better cost-effectiveness.

Ten years after the Barschall study, not-for-profit publishers again emerged in general as a better bargain in terms of cost/impact than commercial publishers. Commercial journals scored worse than not-for-profit journals in all three disciplines: 14.61 versus 8.23 in physics, 42.62 versus 11.55 in economics and 8.69 versus 0.64 in neuroscience.

Nature was not included in the study. But data also gathered by Wisconsin library on the use of thousands of its journals show *Nature* to be one of the most cost-effective journals in terms of 'cost per use' (<http://www.wisc.edu/wendt/journals/costben/stee8.pdf>).

Nature's average cost-per-use was 62 cents last year, and as low as 18 cents at one library. In contrast, over 600 journals had cost-per-uses of more than \$45, with some as high as \$700.

Declan Butler

Bombing of embassy bolsters support for science in China...

Beijing

The efforts of China's research community to argue for policies that favour technological innovation have received a boost from an unexpected quarter — NATO's bombing of the Chinese embassy in Belgrade, Yugoslavia, last May.

The bombing appears to have led to a greater recognition of the importance of science and technology by the Chinese people, including top government officials. The development of cutting-edge technologies is regarded as crucial to national security. Privately, scientists are saying that the bombing has stimulated the government into increasing its attention to science and technology.

Last week, 23 scientists were awarded gold medals — some posthumously — for their contributions to the development of China's first nuclear weapons, missiles and first satellite. The medals were presented by president Jiang Zemin at a ceremony in the Great Hall of the People in Beijing.

The ceremony was presided over by Li Peng, chairman of the National People's

Congress. Jiang and prime minister Zhu Rongji addressed the meeting, speaking highly of the achievements of the scientists. All Politburo members and top government officials were present.

"Without national defence, there will be no nation; without a national defence industry, there will be no national defence," said Zhu recently. A statement issued by the government says: "Without grasping state-of-the-art technologies, we will be bullied."

In the weeks after the bombing, discussions in the science community emphasized the achievements of Chinese scientists in developing the nation's first atomic and hydrogen bombs, and missiles, in the 1960s. Participants stressed the need for China to develop its own high technologies so that the United States could no longer bully it.

The bombing appears to have increased the respect given to scientists. Leading scientists from military research organizations have been invited to talk on television about the technologies used by NATO in its bombing of the embassy.

Tian Xuewen

... as government sets sights on an 'innovation economy'

Beijing

China has confirmed its commitment to encouraging closer links between research institutions and industry, and has instructed all government departments to draw up plans for helping to create an innovation-based economy.

Political bodies have called for greater rewards for scientists and engineers responsible for successful innovations. They will now be allowed to hold shares in companies that exploit their inventions commercially.

The endorsement of the importance of high technology to economic prosperity came during a four-day meeting at the end of last month in Beijing, chaired by prime minister Zhu Rongji and attended by president Jiang Zemin and members of the Politburo, the top authority in China's power hierarchy.

The conference was held to increase awareness among party and government officials of the need to boost technological innovation. A statement issued by the party central committee and the state council before the meeting lists various steps to achieve this. These include accelerating the technological upgrading of traditional industries and increasing the knowledge content of the service sector.

Following the meeting, all major govern-



Great leap forward: the Chinese public is being encouraged to embrace high technology.

ment agencies are discussing how to put these principles into practice. "We will come out with an action plan very soon," says a spokesman for the science ministry.

Science minister Zhu Lilan said the government wanted industries to become the main force in promoting innovation, and that mechanisms should be set up for using venture capital to help achieve this.

Xu Guanhua, vice-minister of the science ministry, said in a television interview that there was inadequate interaction between research institutions and industry in China. As a result, scientific findings were often left unexploited for years. This situation will improve in the future, said Xu, as technologi-

cal brokers will be given a greater role in bringing research institutions and industry closer.

Xu, an expert on remote sensing and former vice-president of the Chinese Academy of Sciences, also expressed concern about the low level of state support for research and development. In 1997, China's spending on research and development was only 0.64 per cent of its gross domestic product. In 1998, the total spending in China was less than the increase in spending alone in the United States over the previous year.

The Chinese Academy of Sciences, China's major research organization, was also due to produce an action plan, to be decided by a top-level meeting on 9 September. But no details have been released yet.

The academy was traditionally focused on basic science, but has been dedicating increasing efforts to applied science and industrial research in recent years. It was selected by the government a year ago to be a target for reform under the National Knowledge Innovation Campaign. More radical reforms of the academy are expected to follow from the conference.

Science parks and high-technology industrial development zones will attract more attention as a result of the conference, according to a paper from the science ministry. T.X.

Japan boosts proteomics and cell biology...

Tokyo

Japan's budget for science and technology is to be substantially increased in the next fiscal year, thanks primarily to a government move to promote industries in areas such as biotechnology, information sciences and the environment.

Particularly marked increases are aimed at developing research into genomics and its applications and into cell biology, which will receive a 12-fold increase in support for the construction of a national research centre.

The government's budget request for the financial year 2000, which begins on 1 April, was unveiled last week. It contains funding for various projects supported by prime minister Keizo Obuchi's Millennium Project (see *Nature* 401, 3; 1999), including a proteomics programme to determine the function of proteins through genome sequencing, and a plan to map single-nucleotide polymorphisms in the Japanese population.

Obuchi's project would be funded by a one-off appropriation to develop an infrastructure for telecommunications, science, technology and the environment for the next century. It is expected to total ¥301.5 billion (US\$2.87 billion), and is intended to help Japan establish a sounder economic base.

This will be added to a relatively modest budget request for 2000, which, at ¥3.18 trillion, is an increase of only 0.8 per cent over this year's science and technology budget.

The budget for genomics, one of the largest beneficiaries of the appropriation,

Table 1 **Highlights of requested budget for science and technology for fiscal year 2000.**

	Request (¥ billion)	% Change since 1999
Total budget	3183*	+0.8
Special appropriation for economic revival	301.5	----
Genome science	81.5	+205
Brain research	30.1	+13.6
Cell biology	12.0	+1233
Life sciences	31.2	+16.0
Climate research	30.2	+314
Industry/government/ university collaboration	51.4	+32.8

*Excluding special appropriation for economic revival

will more than triple to ¥81.5 billion. This includes ¥7.4 billion for the rice genome-sequencing project, which is being accelerated (see *Nature* 401, 102; 1999).

Increased support for genomics is in line with the government's plan to double spending on biotechnology-related research over the next five years (see *Nature* 400, 389; 1999). It is to inject ¥2 trillion, bringing spending on genomics to at least 0.2 per cent of the country's gross domestic product.

The government plans to launch genome projects that can be carried out at an inter-ministerial level. Five ministries, including the Science and Technology Agency (STA) and the Ministry of International Trade and Industry (MITI), will be involved. Such efforts have previously been carried out by individual ministries.

The cell-biology research budget is to rise by a factor of 12 to ¥12 billion, of which ¥10.5 billion is from the special appropriation.

The money will fund research into the development, differentiation and regeneration of plant and animal cells at the new national research centre, due to be built by 2001, and at two centres for applied genomics and plant genome research.

According to Mitsuru Miyata, editor of the Japanese biotechnology newsletter *Nikkei Biotechnology*, cell biology was proposed as a priority area in the life sciences in 1997 by Japan's Science and Technology Council, its principal science policy-making body.

"The government has realized the importance of more integrated research into the interactions of genes and cells as a step up from its previous research in molecular biology," says Miyata, adding that he hopes "it will mark a turning point in Japan's biotechnology research".

Increased support for collaboration between industry, government-run research institutes and universities is another main feature of the budget request, particularly with the impending launch of a new technology-transfer bill, based on the 1980 US Bayh-Dole legislation.

The requested budget for 2000 may still not be enough to meet its target of increasing the country's science spending to ¥17 trillion by 2001, however — overall science spending is expected to total ¥16.7 trillion during the fiscal year 2000.

Asako Saegusa

... while France gives more power to the centre

Paris

French science is to receive a 1.3 per cent increase in funding next year under government budget figures released last week. But this includes a 77 per cent increase in funds directly available to the Ministry of Research for supporting basic research.

This large increase reflects the ministry's desire to increase its control over research spending, and comes as little surprise to those who have been following science minister Claude Allègre's two years in office.

Last year, in an attempt to stimulate research in areas of emerging importance such as genomics, Allègre created the National Science Funds, a centralized fund within the Ministry of Research. Money for this fund rises from FF318 million (US\$50 million) to FF565 million for the 2000 budget. This is earmarked for life sciences, Earth sciences and environmental research.

The overall civil budget for research and development for 2000 is expected to increase



Allègre: more control of strategic spending.

to FF54.6 billion, compared with FF53.9 billion in 1999.

The extra money for the ministry has met with a mixed response. "It doesn't bother me if the government has control over the research," says Pierre Cohen, a member of the National Assembly who headed a

parliamentary commission on French research reforms earlier this year. "But we have always said that it should be transparent and based on principles held by the scientific community."

But unions representing researchers are less enthusiastic. "The government is taking management into its own hands," says Jacques Fossay, a member of SNCS, scientists' main trade union. "There is no

input from the scientific community."

While the budget may launch some of the reforms Allègre is enthusiastic about, such as boosting genomics research, there is no money for other promised reforms, such as creating more posts for young researchers.

"There is little creation of employment," says Pierre Papon, a former head of the country's main research agency, the Centre National de la Recherche Scientifique.

The Funds for Technology Research, aimed at boosting research in high-technology areas, will receive FF905 million, a 35 per cent increase. The money is for research into information and communications technology and for incubating start-up technology companies.

Spending on research laboratories in agencies such as the Centre National de la Recherche Scientifique and the health research agency Inserm will increase by 3.5 per cent, and universities will receive an extra 3.1 per cent for research.

Heather McCabe

US Senate restores proposed cut in science budgets

Washington

The US Senate has restored cuts proposed by the House of Representatives to next year's budget for the space agency NASA and the National Science Foundation (NSF) (see *Nature* 401, 103; 1999).

Last week, the Senate appropriations committee passed an appropriations bill for Veterans Affairs, Housing, Urban Development (VA-HUD) and independent agencies that would fund both agencies at the levels proposed by President Bill Clinton in February. At these levels, the NSF's budget would increase by \$250 million to \$3.9 billion, and NASA's would be held at \$13.6 billion.

The committee's action, which was expected to be endorsed by the full Senate this week, increases the probability that the two agencies will be funded at these levels. The Senate and House proposals will be reconciled later this month, before a joint VA-HUD bill is sent to Clinton for his signature. His advisers have said that he would veto the House's proposed \$1 billion cut at NASA.

NSF director Rita Colwell, whose initiatives in biocomplexity and supercomputer research were threatened by the cuts, says: "We aren't at the finish line yet, but I am breathing more easily".

The Senate obtained extra money for its VA-HUD bill by transferring several billion dollars from the bill for labour, health and education. Both houses of Congress have been tapping the labour bill's allocation all summer, so that it is some \$20 billion short of the \$90 billion its programmes are expected to cost.

According to government officials, Senate appropriators believed the labour bill's allocation was already so low that it had no chance of passing. "Taking another few billion dollars from it won't make any difference," explains one official. The programmes in the labour bill, which include some for the National Institutes of Health, are almost certain to receive temporary funding.

There are various schemes afoot to rescue the labour bill. Most involve using money due to be spent after 1 October 2000, when the financial year ends, so that Congress can claim it did not break its spending cap.

Another option would be for either Congress or Clinton to admit that they cannot operate under the spending caps set by the Budget Balancing Act of 1997, which assumed cuts that Washington has been unable to implement. **Colin Macilwain**

Survey finds deep insularity among Japanese scientists

Tokyo

Japanese researchers should place greater emphasis on the quality of their scientific papers rather than their quantity, and make more effort to publish in international journals and to communicate with foreign researchers. These are the conclusions of a survey released last week by the Science and Technology Agency (STA).

Although 63.5 per cent of the 416 researchers surveyed said that they prefer to submit papers to journals with high impact factors, 26.2 per cent do not consider impact factors as important. This reflects the fact that some institutes — particularly national laboratories and universities — still consider the total output of papers as the main criterion for assessing research activities.

The report also shows that only about 40 per cent of Japanese researchers communicate with foreign researchers using electronic mail or faxes — and a similar proportion appear not to communicate with foreign researchers at all.

The report is based on a survey of researchers at universities and private and government institutes in Japan who produced scientific papers between March 1998 and March 1999. Those who participated in the survey produced an average of two papers during this period, with university researchers tending to produce more papers than those at private institutes, according to the report.

Japan produced more than 10 per cent of the total output of scientific papers in the 1997 index of the Institute for Scientific Information, ranking second after the United States. But a separate analysis based on average citations per paper placed Japan seventeenth, and below the world average.

The STA report raises concern over the number of Japanese researchers who do not publish in international journals, and says that more active participation in international publishing should be encouraged.

The survey revealed that about 40 per cent of the researchers only submit papers to Japanese publications, while 34.1 per cent also submit to international journals.

The survey also shows that only 20 per cent communicate regularly with overseas researchers. Of these, 39 per cent use e-mail, faxes and letters, while 24 per cent say they only communicate in person, and only 0.2 per cent do so by telephone.

"Considering the advances in information technology, I am surprised that more researchers do not use e-mail," says one STA official. "I cannot believe that some still depend on direct person-to-person contact as a means of communicating with overseas researchers."

The report says that Japanese researchers tend to be poor at communicating with overseas researchers. This is one reason for their relative lack of involvement in international science.

Asako Saegusa

Celebratory pictures from Subaru

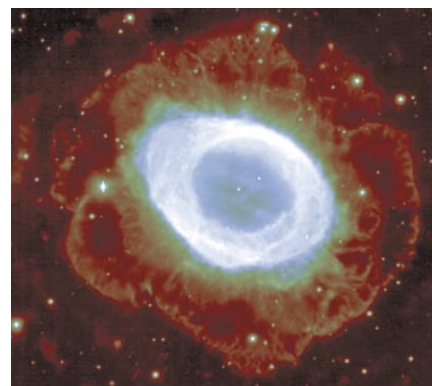
Tokyo

The latest images from Subaru, one of the world's largest optical telescopes, were unveiled last week by Japan's National Astronomical Observatory to celebrate the completion of the 8.3-metre telescope.

Vivid images of the Ring Nebula (M57), the radio galaxy B3 0731 438 and Subaru deep field, a wide field near the North Galactic Pole, were released during the ceremony, which took place on Mount Mauna Kea in Hawaii, where the telescope is located.

Subaru gathered its first light in February (see *Nature* 397, 380; 1999), but last week's images are some of the first to be released since its full operation began in April this year. The images of the Ring Nebula (see right) show the details of its outer halo, revealing its two components: a bright inner part with many loops, and a fainter, more detached, outer part.

These observations contribute to the



Gas rings: the Subaru telescope has captured the inner and outer haloes of the Ring Nebula.

study of how the planetary nebula emerged. Subaru has also captured images of the active galactic nucleus in a radio galaxy 9.2 billion light years from the Earth.

A. S.

US warms to carbon sequestration research

San Diego

The US Department of Energy (DoE) is investigating whether the sequestration of atmospheric carbon dioxide in oceanic or terrestrial ecosystems might be an effective way of reducing global warming.

The department has given \$9 million for three years to two research centres — one each to study the ocean and land — involving scientists from various institutions. Earlier this month, the DoE solicited proposals for \$18 million in carbon sequestration research, with another solicitation expected soon (see www.fetc.doe.gov/business/solicit).

The budding research initiative — which will eventually require far more funds — is drawing together scientists from a variety of disciplines to study cost-effective and environmentally safe methods for the separation, capture and sequestration of the carbon dioxide emitted from fossil fuel. For instance, it might be possible to draw carbon dioxide from the emissions of a power plant, treat it and inject it deep into the ocean for storage.

The research projects are developing in a complex political climate where there is a broad consensus on the need to reduce emissions of greenhouse gases, but varying views on the most cost-effective way to do this.

Some nations, such as Japan and Norway, have aggressively pursued carbon sequestration research. But the United States has funded little research in this area, partly because of political sensitivities. Industry, which depends on fossil fuels, and its congressional advocates feared added costs, while some environmentalists felt carbon sequestration would not curb the burning of fossil fuels.

Last week, the DoE held a workshop bringing together scientists, industry, environmentalists and government officials to seek guidance for “a research road map for practical carbon sequestration technologies”, and to seek comment on its recently released draft white paper, *Carbon Sequestration — State of the Science*.

Scientists attending the two-day workshop in Maryland described it as an encouraging exercise that showed new cooperation among disparate parties. But they also said that a shortage of data means it will be decades before some new forms of sequestration find large-scale applications.

“There is so much we don’t know,” says Howard Herzog, a chemical engineer at the Massachusetts Institute of Technology’s energy laboratory, “but if we don’t do the groundwork now, we won’t be ready”.

With terrestrial sequestration, carbon dioxide emissions would be pumped into geological formations or old oil wells, reduced by farming techniques (such as eliminating deep ploughing), or offset by forestry procedures such as planting more



Deep thought: the effects of liquefied carbon dioxide on the ocean environment are already being studied by the University of Hawaii.

trees. Small-scale applications of these methods are already being applied in some countries, mostly without controversy.

In contrast, ocean storage of carbon dioxide is seen as having great potential because of the vast areas available; but environmental issues make it a sensitive alternative. Injecting carbon dioxide may change the ocean chemistry, possibly with harmful ecological consequences.

It is these aspects of carbon storage that scientists want to study. Next month, the UK-based Environment Council will hold a workshop in New York to explore whether the injection of carbon dioxide into the ocean is “a robust environmental management strategy”. In May, the organization held a similar workshop in London.

With ocean sequestration, carbon dioxide would be liquefied and then pumped into the depths, where huge amounts of carbon already exist in forms such as carbonic acid or carbonate.

Dispersing carbon dioxide in the ocean is also being considered, but this may be even more controversial, because of concerns

about more widespread ecological damage.

“We are being quite cautious,” says Jim Bishop, a marine chemist at the Lawrence Berkeley National Laboratory who is a co-director of the DoE’s new Center for Research on Ocean Carbon Sequestration. “Before we pump carbon dioxide into the sea, we need to better understand how it will work”.

To that end, a team of scientists in Hawaii is preparing to conduct the most extensive experiment to date on injecting liquefied carbon dioxide into the deep ocean. The \$5-million, four-year project is being funded mostly by Japan, with assistance from the United States, Norway and Canada. Australia is also expected to join.

Stephen Masutani, a mechanical engineer at the University of Hawaii who is one of the project’s leaders, says the plan is to inject liquefied carbon dioxide 900 metres deep off the Kona coast in the summer of 2001.

The injection will use a small, removable pipe laid into the ocean from the Hawaii Natural Energy Laboratory. A remotely operated, mobile platform will monitor carbon dioxide dispersal at the injection site. “The idea is to obtain data for predictive models,” said Masutani. “We want to see what the changes in the sea chemistry are.”

But, despite more than two years of planning and laboratory work, Masutani says the experiment is under scrutiny from local environmental groups concerned about the impact on sea life and water chemistry.

The Hawaii team is spending considerable funds on explaining the experiment. If a full environmental impact statement is required, says Masutani, it would delay the experiment beyond the funding period set by Japan, possibly scuppering it. Decisions on the environmental assessment needed are expected in the next few months. **Rex Dalton**

German museums face cuts

Munich

Publicly funded research at Germany’s six largest science and history museums — including the Deutsches Museum in Munich — is to be cut next year by at least 7.5 per cent.

Areas likely to be immediately affected at the Munich museum include its efforts to promote the public understanding of science, for which it has recently taken on responsibility for nationwide coordination.

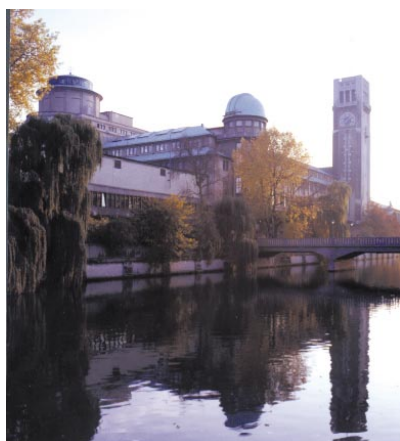
Some of the biological research at the Zoological Research Institute and Museum Alexander Koenig in Bonn would be “paralysed” by the threatened cuts, according to Michael Schmitt, a biological

researcher there. “It would mean the end of my research project on the phylogeny of *Chrysomelidae* [leaf beetles],” he says.

The museums’ research departments form part of the Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) — formerly called the ‘blue list’ — a loose association of 82 research and service institutes that makes up Germany’s ‘fourth pillar’ of non-university research.

The WGL is funded equally by the federal government and Germany’s 16 Länder (states), as are the Max Planck Society (MPS) and the university funding organization, the Deutsche Forschungsgemeinschaft (DFG).

The MPS and DFG will see their budgets



Shrinking asset: Munich's Deutsches Museum.

increased next year. But the science museums are not being exempted from an overall decrease of DM30 billion (\$16 million) in public spending imposed by finance minister Hans Eichel to reduce the national debt.

"This is a clear [act of] discrimination against the WGL, and science museums in particular," says an official of the Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), which coordinates regional and federal research policies.

The WGL already faces the problem that responsibility for funding its institutes is spread over ten ministries. As for the science museums, funding responsibility shifted last year from the Ministry of the Interior to the new Ministry of Cultural and Media Affairs.

Culture minister Michael Naumann has announced that the cuts will be imposed equally across all spending areas of his ministry. With a relatively small portfolio of DM1.7 billion, Naumann says that he cannot prioritize some areas by excluding them from the cuts. In contrast, the research ministry has been able to respond more flexibly by shifting its priorities.

Research costing DM46.2 million is being carried out at the six science museums this year. Germany's science council, the Wissenschaftsrat, has confirmed the high quality of the museums' research departments, and recommended that the proportion of research to other museum activities should be maintained.

But the 7.5 per cent cut will hit the museums hard. "The cuts come at a time when we had been about to expand our activities," says Helmuth Trischler, director of science and vice-director-general of the Deutsches Museum. "As a consequence, we will not be able to do what we intended to." **Quirin Schiermeier**

US university finds use for Air Force research reactor

San Diego

The University of California (UC) at Davis received approval last week to acquire a research nuclear reactor, up to now used by the US Air Force (USAF), for use in studies in engineering, materials science, agriculture and medical therapies.

The reactor, located at the USAF's McClellan Air Force Base, about 20 miles from the university, was built in 1990, making it the youngest of the TRIGA (Training, Research, Isotopes, General Atomics) class of research reactors which are no longer being built in the United States.

The reactor will form the core of a developing research programme at Davis, which officials say will become more important in future years as older research reactors are decommissioned and not replaced.

"Eventually, UC Davis may be the only show around" for research reactors, says Kevin M. Smith, UC Davis' vice-chancellor for research. Officials say that research reactors are difficult to build because of increasing costs, the extensive environmental assessment process required and community opposition.

Under terms of the agreement approved last week by the UC Board of Regents, the reactor must be used at least 51 per cent of the time for research; the remainder of the time it can be used for revenue-producing commercial work, which university officials hope will cover some of its operational costs.

The Air Force has been using the reactor for neutron radiography to detect hidden cracks and defects in aircraft wings. Some

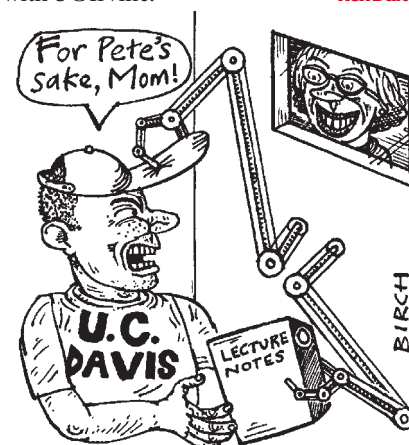
similar work may continue as UC Davis builds up its reactor-based research programme.

After a competitive selection process, the university has appointed Science Applications International Corp. (SAIC) of San Diego as the reactor's operator. Most employees now working at the site will be hired by SAIC.

The US Department of Energy is providing \$8 million to fund research and operational costs for the reactor for the next four years. And the federal government is giving \$17.6 million for the future decommissioning of the reactor, expected in about 30 years.

As UC Davis develops a programme of teaching and research around its reactor, officials plan to collaborate with UC Berkeley's department of nuclear engineering, and with UC Irvine.

Rex Dalton



UK universities look to enterprise

Sheffield

Eight UK universities are to set up 'centres of enterprise' to exploit their research and transfer ideas to the private sector. Steven Byers, the Secretary of State for Trade and Industry, announced the news last week at the British Association for the Advancement of Science's annual Festival of Science.

One of the responsibilities of the centres, costing a total of £25 million (\$40.5 million), will be to incorporate the teaching of enterprise into the universities' curricula, equipping scientists and engineers with entrepreneurial skills. The government hopes the centres will play a key role in regional development, and wants them to link up with local companies.

Byers also announced the extension of

the network of Faraday Partnerships, which aim to link science and business in specific areas of technology by bringing together consortia of independent research organizations, universities, businesses and financiers. Four new partnerships, with a budget of £8.8 million, were announced. Consortia are invited to propose new topics in their bids. "We want your ideas, we want to see partnerships strengthening industry clusters in their localities," said Byers.

Byers announced additional funding for the four existing Faraday Partnerships. He said this was because the previous government had underfunded the centres, and added that the extra support would allow them to restructure their 'industry-facing' activities.

Natasha Loder

Portugal's science plots its future

Research in Portugal has grown apace over the past decade, and the government is reviewing its achievements, problems and prospects.

Portugal's Ministry of Science and Technology (MST) has launched a nationwide consultation on how the country's science and technology should develop over the next seven years.

The consultation is based on a 'white paper', made up of a series of working documents. The government is seeking comments and proposals from public and private research centres, and from groups such as public authorities and private industry.

The ministry has set up a web-based Forum on Scientific Policy to conduct consultation. The website includes a letter to Portuguese scientists from the research minister, former physicist José Mariano Gago, explaining the country's short- and medium-term research objectives.

Gago hopes to stimulate debates that will produce information on the development and resources of different disciplines. The Observatory of Science and Technology (OST), a body based at the MST, will coordinate the collection of data on the nation's science and technology.

Cooperating with Europe

Based on the results, the MST will complete an Integrated Programme of Science and Technology, review the implementation of the 1988 Law of Scientific and Technological Development, and prepare the next European Commission programme of support for science and technology.

Gago says that the white paper's main conclusions already form the basis of the new Scientific and Technological Development Programme (2000–2006), which has been approved by the government and will be presented to the commission shortly.

It includes two interrelated operational programmes: the first, called Science Technology Innovation, has a budget of 192 billion escudos (US\$1 billion); half of this will come from the Portuguese government and half from the European Union. The second, Information Society, has a budget of around 140 billion escudos and will be funded from the same sources.

The provisional versions of these programmes were made public at the end of July. Gago says that they will be formally submitted to Brussels within a few weeks.

The white paper includes a report, put together by the OST, which describes Portugal's science policy over the past few years and its future objectives in research and technology, including a special section devoted to information technology.



On the up: the Universidade Nova de Lisboa. Gago (inset) is consulting on science policy.

The report also analyses the international activities of Portuguese scientists. In 1996, 49 per cent of the publications with Portuguese coauthors appearing in international scientific journals were international collaborations, up from 28 per cent in 1980–81.

The white paper includes 'research profiles' covering physics, mathematics, chemistry, health, engineering and environmental science, produced by Portuguese specialists nominated by the ministry. These experts are the same as those who coordinated 20 international evaluation panels that assessed the performance of publicly funded research units three years ago (see *Nature* **387**, 115; 1997).

The earlier evaluation consisted of an analysis of the resources and achievements of individual research units, visits to the units and a final report with recommendations for each field.

The reports assessed, with the aid of the OST, the scientific potential of each field and gave information on research activities, human, technological and financial resources and the strengths and weaknesses of research units.

The white paper 'can help show the weak points in the system'

Most of the experts involved agree that the panels were useful, rigorous and fair, although some suggested there should have been more experts from outside the European Union, especially from the United States.

But Gago says that some scientists told the ministry that the proportion of US experts was too high, and that they should be replaced by Europeans. The evaluation will be repeated every three years — the next is being carried out over this month and October.

Careers and communication

Carolino Monteiro, a human geneticist at the Universidade Nova de Lisboa, says that the white paper is "a very important incentive for the Portuguese scientific community, as people can start to talk and give their opinion".

Edgar da Cruz e Silva, director of the Centre for Cell Biology at the Universidade de Aveiro, thinks the white paper is "a very interesting exercise that can help identify the weak points of the system and sort problems out, not only in a given research unit but also externally".

The move reflects the MST's efforts, led by Gago since 1995, to reform Portugal's scientific system. Already there has been a ten per cent annual average growth in the number of PhDs, a 14–16 per cent annual growth in the public research and development budget since 1995, and a trebling in the number of Portuguese publications cited in the Science Citation Index over the past decade.

Despite this progress, some of the contributions to the white paper warn of a lack of coordination of research in Portuguese universities. Cruz e Silva says that in 1995 his university set up an 'institute of research' made up of all the research units.

"So far, the experience shows that such an institute has been very useful in managing and coordinating science and technology activities in the Universidade de Aveiro," he says.

Monteiro is worried that many Portuguese scientists cannot find research posts after their postgraduate or postdoctoral training. The white paper's reports — supplemented by the coordinators of the evaluation — reflect these concerns.

The report concludes: "Appropriate recruitment policies appear to be urgently needed to achieve not only a sustained expansion of the system, but also progressively younger research personnel."

Xavier Bosch

Healthy economy spurs new UK review of science spending

London The UK government has told the research councils to bring forward their next priority-setting exercise, as part of a broad review of public spending that is to take place at least six months earlier than expected.

The move has been prompted by the Treasury's desire to re-examine its spending plans in the light of a strong economy and a lower than expected rate of inflation. Some see it as an opportunity to bid for extra funds for areas of science — such as European physical sciences projects — that did relatively badly in the last spending round.

But, while there is optimism in the research councils at this prospect, there is also awareness that it will take a considerable effort to ensure that science does as well as it did in last year's Comprehensive Spending Review, which resulted in an increase in the councils' budget of £700 million (US\$1.14 billion) over the three years up to 2001.

Biodiversity survey in US reveals mounting problems

Washington US ecologists are hoping that a new, comprehensive survey of the country's

biological resources released last week will help policy-makers understand the pressures on ecosystems — and persuade them to provide more money for studies of biodiversity.

The survey brought together more than 200 authors to conduct a detailed, region-by-region audit of the country's ecosystems and the problems that confront them. According to Michael Mac, head of fisheries research at the biological resources division of the US Geological Survey, and lead author of the 1,000-page document, these problems continue to mount. "In summary, it's all bad news," he says.

Computer musician strikes golden note

Paris This year's Médaille d'Or of France's national research agency, the Centre National de la Recherche Scientifique, has been awarded to Jean-Claude Risset, a scientist who specializes in computer-generated music. During a career that began in the mid-1960s, Risset has focused on simulating instruments electronically and studying perceptions of sound.

At the Media Lab at the Massachusetts Institute of Technology, Risset developed a program that allows a computer to respond to a pianist and act as a real-time accompanist. Risset has also composed and

recorded works that combine instruments, voice and computer-generated sounds.

US and South Africa in pact over AIDS drug prices

Cape Town The United States and South Africa have come to a verbal agreement over South African legislation which the United States believes would allow the South African health minister to violate international principles of patent protection to procure AIDS drugs and other medicines at reduced prices.

In a statement released in Washington last week, US trade representative Charlene Barshefsky said the United States "very much appreciated South Africa's assurance that, as it moves vigorously forward to bring improved health care to its citizens, it will do so in a manner consistent with international commitments and that protects intellectual property rights".

Rush for funds for industrial research

Munich Germany's federal research ministry has been overwhelmed by nearly 500 applications for its InnoRegio competition, an initiative launched earlier this year by Edelgard Bulmahn, the federal research minister, to promote cooperation between research

institutes and small and medium-sized companies in east Germany.

Despite the unexpectedly high response, the jury of independent experts will meet its November deadline for selecting 25 projects for the next round, says Michael Catenhusen, state secretary for research. Each will receive DM300,000 (US\$160,000) to develop their proposals. The winners, to be selected next summer, will share DM500 million of federal funding over a five-year period. The initiative is modelled on the successful BioRegio competition, which Bulmahn's predecessor, Jürgen Rüttgers, launched in 1995 (see *Nature* **384**, 298; 1996).

Arizona and Australia look out for asteroids

San Diego The University of Arizona and the Australian National University are collaborating to upgrade and operate a telescope in Australia to provide a much-needed asteroid survey in the Southern Hemisphere. The telescope, at the Siding Spring Observatory near Coonabarabran, New South Wales, will complement the surveys of near-Earth objects being carried out at five sites in the United States.

The US space agency NASA, which is funding the observation sites, is expected to provide additional funding later this year to

computerize the Schmidt-style telescope at Siding Spring. Arizona has contracted the Australian astronomer Robert H. McNaught to provide observations of potentially hazardous asteroids.

US universities pay up in 'improper bills' suit

San Diego The University of California at San Diego (UCSD) and the University of Washington in Seattle agreed last week to pay a total of \$8.3 million to settle a federal lawsuit that alleged they had improperly billed government health programmes for experimental heart devices.

UCSD will pay \$4.7 million and Washington \$3.6 million to settle the lawsuit, filed by a former medical equipment salesman. The universities had argued that they were trying to provide the best treatment for patients under rules that were ambiguous. The Department of Justice is continuing to investigate similar allegations against more than 100 other medical research institutions.

Welcome to the undersea world of designer fish

Boston The Museum of Science in Boston has launched a 'virtual fish tank' that



Fishy tale: design-a-fish program will enhance understanding of behaviour, says museum.

demonstrates how simple rules can create complex behaviour patterns in nature. The simulated undersea world (above), which measures 160 square metres, lets visitors design and launch their own fish — and watch the behaviour patterns that emerge.

The museum says that the tank is a compelling example of the value of computer simulations and high technology in understanding science. The FishTank, first launched in June 1998 at another site, combines 3D computer graphics and real-time interactive character animation, and now features a side view of the aquarium.

It is hoped that the exhibit will attract more than 1.8 million visitors a year.

Taxonomists are an endangered species in Europe

Sir — Studying and making an inventory of the planet's vanishing biodiversity are apparently not of great concern to European decision-makers. Taxonomists face increasing difficulty in finding work and funding. Many European universities no longer offer taxonomy courses. It is getting harder to maintain important specimen collections and handle loans.

There has been a catastrophic reduction in cataloguing the still largely undescribed tropical biota — the West's responsibility, as there are few trained taxonomists in tropical countries. Compiling an inventory of the tens of thousands of larger fungi in tropical Africa depends on three or four European mycologists. And this at a time when the scientific community has started to realize how much fungal diversity has been underestimated in the past — barely 10% of organisms have been described.

Several French and Dutch research centres with international reputations in taxonomic mycology have been dramatically cut in size. In Britain, retiring taxonomists at the botanic gardens of Kew and Edinburgh have not been replaced, and the International Mycological Institute has ceased its independent existence. As far as the taxonomy of most of the typical larger mushrooms is concerned, there is only one professional mycologist left in the Netherlands, two in France, and none in Britain. Switzerland looks likely to have none soon if Zurich follows the example of the University of Lausanne and fails to replace departing staff. Retiring mycologists are often heads of laboratories or research units, and their retirement usually puts an end to taxonomic mycology there.

In contrast, US institutions such as the Missouri Botanical Garden are dynamic and growing fast. With a policy of permanent presence on all continents and scientists at Europe's most famous institutions, US efficiency is imposing itself on one of the richest legacies of European history.

Bart Buyck

Muséum National d'Histoire Naturelle, Laboratoire de Cryptogamie, 12 rue Buffon, 75005 Paris, France

Don't leave the biology out of bioinformatics

Sir — Marlie MacLean and Colin Miles highlight the skills gap for 'bioinformaticians' in the United Kingdom and elsewhere (*Nature* **401**, 10; 1999). There is little doubt that, given the resources, it will be possible to find many people capable of

manipulating the enormous volumes of data being generated, but such people must have a knowledge of and genuine interest in biology. The solution is either to identify those who are adept at both computer science and biology, which will be difficult, or to create interdisciplinary teams, which is a much more realistic approach.

The present headlong rush to expand bioinformatics could lead to the creation of isolated groups that have little interaction with molecular biologists and biochemists. If this occurs, the subject will fail to achieve the promise that we are led to expect in terms of a better understanding of biology and disease.

Peter Campbell

Department of Biochemistry and Molecular Biology, University College London, London WC1E 6BT, UK

Courage could win back confidence in science

Sir — Bradford Hawkins recently described an increase in the number of technical and trivial errors appearing in the published scientific literature (*Nature* **400**, 498; 1999). This, it was proposed, is a consequence of the 'publish or perish' philosophy of academic institutions. I would like to draw attention to an equally destructive trend: that of publishing banal descriptive reports that fail to enthuse the reader, or indeed the author.

The educational path to becoming a research scientist teaches the importance of detail, experimental rigour and assimilation of large amounts of scientific fact. New graduates are efficient at compiling literature searches, conducting experiments and analysing data. They are often eager to ask and answer fundamental questions of science and fit their experimental findings into a wider perspective of principle. However, they soon find that the scientific community does not reward them for this approach. Until recently, funding bodies and scientific publishers (presumably on advice from their reviewers) have executed a precautionary principle, favouring a conservative approach in scientific proposal and interpretation. The result is a failure of confidence to hypothesize.

The need for imaginative science is at last being recognized by major funding bodies — the W. M. Keck Foundation, among others — that provide funding for young investigators to establish more creative research projects. But, so far, this philosophy is rarely supported in the publication of scientific reports. It is more difficult to get a speculative comment through a reviewer than it is to publish several megabases of genetic sequence. Max Perutz recently drew attention to the plight of too many young post-docs with diminishing career prospects

(*Nature* **399**, 299–301; 1999). This indirectly encourages predictable research and publication of trite interpretations of data instead of new insights, because, understandably, scientists are more interested in remaining employed than in taking the risk of rejection from reviewers.

There is an urgent need for the responsible communication of scientific discovery to counteract the public's loss of faith in science. While being didactic in form, it should convey a sense of the excitement and momentum of scientific discovery without being sensationalist. What better place to start than in the scientific press? For this to happen, publishers, reviewers and scientists need encouragement to be brave.

Susan Daenke

Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

Sprucing up one's impact factor

Sir — Journal impact factors (IFs) were introduced in the 1970s to rank different journals by citation analysis¹. IF is defined as the mean number of citations received in a particular year to articles published in the journal in the preceding two years.

It has been recognized for some time that the absolute values of IF, for a given qualitative journal standard, vary widely among disciplines (for example, biology, chemistry and engineering) and that review (secondary publication) journals have much higher IFs than do the primary publication journals carrying original research articles. Nevertheless, the IF has become an important parameter helping librarians to make difficult decisions concerning journal subscriptions, and consequently for the journals themselves in their attempts to increase sales and advertising revenues.

We wished to determine recent changes in the editorial or publication policies of different journals in biomedical fields that may have a bearing on their IFs. Our search has been limited, but the following findings are noteworthy.

(1) Since the early 1990s, many primary-publication biology journals have introduced 'mini-reviews' or their equivalent. Review articles attract citations more rapidly, and in larger numbers, than primary articles.

(2) Research articles in medical journals are often the subject of short comments that are published, with a response from the original authors, in a subsequent issue. For the first time in its 175-year history, *The Lancet* altered its policy in 1996 so that an explicit (that is, countable) citation to the original article is now included in every one

of the published letters of comment; earlier, such citations were included in the text itself and were not countable.

(3) Perhaps most important, one must consider the case of articles that are not counted as 'source' items (for the denominator) in the calculation of IF. These include editorials, book reviews, letters to the editor, and obituaries. Perhaps surprisingly, citations to such 'non-source' articles are counted in the numerator in the IF calculation. For the journals concerned, such items may therefore be taken to represent instances of 'nothing lost, anything gained'. Included in the examples of non-source articles are the numerous short clinical or laboratory study reports published as Letters to the Editor in *The Lancet*, some of which may even be considered as citation classics; a pair^{2,3} in 1993 on non-01 cholera received approximately 200 citations in the next two years.

Some categories of published items originally classified as source articles have been reclassified as non-source. For example, meeting abstracts published in *FASEB Journal* were reclassified as non-source articles from 1988, and the IF for the journal registered a leap from 0.24 in 1988 to 18.3 in 1989. In 1983, *Nature* started publishing Scientific Correspondence which, together with the prestigious News and Views section, now comprises a large repository of citable non-source articles. Many other journals appear to be following these trends. When one compares the number of pages devoted to source articles with those for non-source items, the ratio for *Nature* had more than halved, from 3.5 in January 1977 to 1.6 twenty years later, even though the total number of pages in the journal was virtually unchanged over this period. If nothing else, our findings support the case for a change in the present method of IF calculation, so that citations to source articles alone are counted.

Finally, we have identified a loophole that could allow a less than scrupulous, perhaps obscure, journal to increase its IF from 0.1 to a healthy 2.1, by the mere expedient of adding two spurious self-citations in each of its source articles. Is it possible that this may be happening already?

J. Gowrishankar, P. Divakar

Centre for Cellular & Molecular Biology,
Uppal Road, Hyderabad 500 007, India

Sir—Scientific institutions are increasingly judged on the quality of the journals in which their staff publish papers. Journal quality is usually presented as its impact factor (IF), the number of citations in a given year to papers published in the two previous years, divided by the number of items published in those previous two years (*Journal Citation Reports*, Institute for Scientific Information, 1997).

As veterinary researchers, we sometimes find ourselves searching for possible human angles in our work, so that we might publish in medical journals, which tend to have significantly higher IFs than their veterinary counterparts. But we have spotted a simpler and more effective approach that will allow us to publish in appropriate places and still get high ratings. As an example, *The Veterinary Record* has an IF of about 1, based on approximately 600 citations and 600 papers published in 1995 and 1996. Our institute publishes about 300 papers in two years. Our director need only instruct us all to cite at least two papers from *The Veterinary Record* in every paper we publish from now on, however loose the connection, for the IF to quickly double. *The Veterinary Record* would move from being in the top 40% of journals to being in the top 15%.

We could have an even greater impact on journals that publish fewer papers. For example, if our director applied this policy to *Veterinary Research Communications*, the IF of that journal would increase from less than 1 to more than 6, moving it into the top 3% of journals. If our institute teamed up with two or three others, we could rapidly create a competitor to *Nature*. Unethical, perhaps, but legal and very much in our interest.

**Matthew Baylis, Michael Gravenor,
Rowland Kao**

Institute for Animal Health, Compton Laboratory,
Compton, Newbury, Berkshire RG20 7NN, UK

1. Garfield, E. *Science* **178**, 471–479 (1972).
2. Ramamurthy, T. et al. *Lancet* **341**, 703–704 (1993).
3. Albert, T. J. et al. *Lancet* **341**, 704 (1993).

No dirty tricks in merger ballot

Sir—Rex Dalton's article about the proposed merger of the Optical Society of America (OSA) and the International Society of Optical Engineers contains a number of misleading statements (*Nature* **400**, 605; 1999).

Dalton refers to accusations that a recent membership mailing of OSA exemplified "inappropriate leadership tactics". In fact, this half-price membership effort is an annual drive that was offered to all non-members in the society's vast database. No one group or specific out-of-house list was targeted for this mailing, and the inference that this solicitation was made to influence the proposed merger vote is without merit.

Dalton's report of threatened legal action against an unnamed OSA scientist is false. A letter sent by OSA's executive director to the scientist clearly states concern over the unauthorized use of a mailing list, but it does not make a legal threat. The society has traditionally taken a respectful stance regarding paper and electronic-mail

intrusion into the lives of our members and always has acted to protect its mailing lists as an asset of OSA.

The article also leaves the mistaken impression that opponents of the merger have somehow been silenced in this debate. The board of directors and the staff of OSA have worked diligently to ensure that accurate, timely and informative materials were (and are) provided to our members so that they can make an educated decision on the proposed unification. There have been open forums at society meetings, links provided to the opponents' website, all-member communications providing the opponents' views mailed at the society's expense, and a ballot package sent to all eligible voters presenting equally arguments for and against the proposal.

This vote is extremely important to the future of our society—and the field of optics—and every effort has been made to keep the information flow fair and balanced.

Elaine Gansz Bobo

Optical Society of America, 2010 Massachusetts
Avenue NW, Washington DC 20036-1023, USA

Call a halt to strong-arm tactics over GM crops

Sir—It is surprising that representatives of an organization that is reported to use acts of destructive force to achieve social and political goals are given a forum in *Nature* on the subject of how to restore public trust in science. I refer to the Commentary by Greenpeace's Benny Haerlin and Doug Parr (*Nature* **400**, 499; 1999). One such act is the destructive attack on fields of genetically modified crops by a group that included the head of Greenpeace UK, as reported by *The New York Times* (23 August).

Press accounts indicate that this was only the latest in a series of attacks by Greenpeace and allied organizations. How can we now tell if a future refusal by farmers to grow genetically modified crops will not really be based on fear disguised as conviction instead of genuine conviction?

European political and religious history is replete with groups that have found the use of apparently peaceful propaganda together with the selective use of brute force to be an effective tool to change public opinion. It is sad that this seems to be happening today in Britain.

Haerlin and Parr suggest that the values of society should be paramount in the debates they discuss. This would have been an interesting suggestion to give to Galileo. Is the use of force in civil discussions one of the values they have in mind?

Manfred Philipp

City University of New York, 250 Bedford Park
Boulevard West, Bronx, New York 10468, USA

Balancing the Earth's accounts

Governments could safeguard the world's biodiversity with a small fraction of the money they spend on environmentally harmful subsidies.

Alexander N. James, Kevin J. Gaston and Andrew Balmford

Despite recent estimates that the Earth's ecological systems are worth about US\$33 trillion annually¹, the comparatively low cost of maintaining the biological diversity that underpins these services is ignored. This cost, which we examine below, continues to go unmet by governments and foreign donors, while the attitude that 'we can't save everything' is used to justify saving far too little. New data on the cost of maintaining and expanding nature reserves, and of 'greening' natural resource sectors such as agriculture, suggest that these costs are not as high as might be thought. In our view, conserving the planet's biodiversity would cost less than a quarter of the amount governments currently spend on environmentally harmful subsidies, recently estimated at \$1 trillion per year^{2,3}. Intermediate conservation goals, such as protecting existing nature reserves, could be accomplished for a small fraction of this amount.

The great conservation bargain

The amount spent on nature reserves by governments and foreign donors — arguably the cornerstone of conservation efforts — is pitiful: only \$6 billion per year⁴. (See Table 1, overleaf, for a breakdown of these and other figures.) This works out to \$453 per km² globally, though just \$93 per km² is spent in the tropical countries. Estimates by us and global-reserve managers⁴ show that adequate budgets for protected areas would cost only an additional \$2.3 billion per year. Favourable cost structures in the developing countries keep this figure low: summing current and shortfall expenditures, overall costs there average \$277 per km², compared with \$1,090 per km² in developed countries.

It would cost more to expand the global system of parks and protected areas to a minimum standard of 10% of the land area of each region⁵ — although perhaps less than one might expect, as this target is already met in some areas⁶, and 90% of the additional area required is in developing countries. Even if a higher standard were set, such as making 10% of the Earth's land area strictly protected (IUCN Categories I–III; ref. 6), plus retaining the existing multiple-use reserves (4.7% of world area), the costs would still be quite manageable.

A full accounting of these expansion costs must include money for biodiversity surveys

to ensure the ecological representativeness of the new reserves, for land purchase, and for administration and management. Surveying 20% of the global land area would require roughly \$4.2 billion — surveys typically cost about 30% of annual management⁷. Buying the land so that 10% of the area of each region was in strictly protected reserves would require approximately \$164 billion, based on recent conservation land purchases⁷ at around 50 times the annual cost of management. These survey and acquisition costs are both one-time expenditures, which if implemented over 30 years would entail an annual outlay of \$10.9 billion (amortized at 5%). Management and administration of these areas would add around \$3.3 billion annually.

In sum, we estimate that to buy and manage a broadly representative system of nature reserves covering nearly 15% of global land area (10% strictly protected) would cost roughly \$16.6 billion per year on top of the \$6.0 billion currently spent.

Compensating the losers

When an area is strictly protected, local communities suffer significant costs; effective long-term conservation will depend upon appropriate compensation⁸. In principle, these opportunity costs are equal to the value of the lost resources, including all intangible benefits. In an undistorted market, land prices fully reflect these values, as landowners will not voluntarily sell for less than full value. Therefore, the opportunity costs of expanding the protected area system will be met in the land prices, as long as they are fair.

However, compensation should also be

paid for the opportunity costs of many areas that are already protected. Exclusion is felt most acutely in Category II, III and IV reserves in developing countries (Category I reserves are mostly uninhabited and Categories V and VI allow substantial resource use⁶). We believe that local communities' property rights to these resources should be recognized and compensation paid accordingly.

If compensation payments are tied to land values in these areas, then the land costs estimated above imply that another \$49.5 billion would be required in compensation. This payment could be structured to provide incentives, possibly in the form of an annual payment equal to the resource value discounted at 10%, \$4.9 billion.

Even though compensating local communities properly for lost access to protected areas is affordable, the challenge lies in distributing the payments fairly, so that they provide incentives for long-term conservation. Community-based conservation programmes — in which incentives for sustainable use of resources in protected areas are created through certain ownership rights — are an alternative solution. But recent experience with such programmes has not been encouraging, and there is evidence that the conservation values of protected areas continue to deteriorate, despite a decade of practice^{9,10}. Rather than expand community-based conservation programmes, we believe local communities could be paid direct compensation for loss of resources, and helped to develop alternative employment opportunities. The nearly \$5 billion per year spent in compensation for opportunity costs would



Cutting edge: as development results in the fragmentation of natural habitats — such as Brazil's Atlantic forest — protecting the environment in the landscape between reserves becomes vital.

A. BALMFORD

support considerable experimentation with these programmes throughout the developing world.

Greening the wider landscape

A truly comprehensive global conservation plan must include the cost of protecting biodiversity in the matrix of landscapes beyond reserves — in the agriculture, forestry, freshwater, coastal and marine sectors. Biodiversity remediation might cost \$34 billion for forestry, \$1 billion for freshwater and \$14 billion for coastal and marine systems, according to Agenda 21 (ref. 11). Biodiversity conservation in the farming sector would cost far more — around \$2.4 billion annually in the United Kingdom alone^{12,13}. The global cost of 'greening' agriculture might amount, very roughly, to \$240 billion, given that the United Kingdom accounts for around 1% of global cereal production¹⁴. Wider costs may thus total about \$290 billion per year.

The cost of maintaining biodiversity in the human-dominated landscape shows what a remarkable bargain nature reserves are. Conservation of an ecologically representative global network of protected areas would cost only \$27.5 billion per year — \$6.0 billion in current expenditures, \$2.3 billion for strengthening the management of existing reserves, \$14.3 billion in expansion costs and \$4.9 billion in compensation. Given the key role of protected areas for conservation in the immediate future, meeting these basic expenditures should be a far higher priority in government budgets and on the international agenda.

Cheaper than the alternative

Our rough estimate of \$300 billion annually for a comprehensive global conservation programme is a relatively small amount when compared with government expenditures on activities that harm the environment. 'Perverse' subsidies to promote agricultural production, energy use, road transportation, water consumption and commercial fishing total between \$950 billion and \$1,450 billion per year^{2,3}. These subsidies keep resource prices below market levels, thus encouraging over-exploitation of the natural environment.

For example, agricultural subsidies, which frequently harm biological diversity by simultaneously driving intensification in some farmed areas and excessive land conversion in others, average \$82,500 per km² in the European Union and \$16,100 per km² in the United States³. By comparison, these countries spend on average less than \$2,000 per km² on their national parks and nature reserves.

Implementing a global plan to conserve biodiversity should go hand-in-hand with the gradual removal of these subsidies. If subsidy reform were linked to investment in environmental protection, a small shift in government spending patterns would

Table 1 **Funding requirements for a global system of protected areas**

	Area (km ²)*	Existing reserves			Additional reserves		
		Expenditure	Shortfall	Compensation†	Area	Survey and purchase‡	Management
North America	3,924,047	3,350	335	0	34,700	180	33
Latin America and Caribbean	2,121,753	216	297	1,243	1,174,150	789	242
Europe	603,601	1,171	505	0	339,393	3,054	943
Russia and CIS	657,935	51	57	268	1,786,605	944	292
Developed East Asia	36,825	453	45	0	33,533	1,471	454
Developing Asia	1,578,644	105	567	1,283	1,068,340	1,439	442
Sub-saharan Africa	2,074,451	245	253	1,896	1,571,258	1,417	437
North Africa and Middle East	1,037,576	43	195	231	1,148,980	850	263
Australia and New Zealand	1,109,024	297	30	0	232,624	227	68
Pacific	13,113	35	7	25	54,307	555	172
World	13,156,969	5,967	2,290	4,947	7,443,891	10,926	3,347

All costs in 1996 US\$ millions. *IUCN Categories I–VI. † Annual payment at 10% of resource value (see text); ‡ Annual cost when amortized over 30 years at 5%. Full details of calculations and data sources are available from the authors.

accomplish major conservation objectives. An effective global system of nature reserves would cost about 2% of the annual expenditure on environmentally harmful subsidies. A truly comprehensive global conservation programme that incorporated biodiversity into all major natural-resource sectors could be launched for only about 20% of the cost of these subsidies.

But harmful subsidies are not necessarily concentrated where biodiversity expenditure needs are greatest. For example, three-quarters of all harmful subsidies are spent by governments of developed nations within their own borders³. Although this may match the geographical spread of funding required for conservation in the wider landscape fairly well, much of the extra spending on protected areas is needed in the developing countries, where perverse subsidies are much smaller. The Convention on Biological Diversity could play a catalytic role in redistributing such funds.

The Convention contains the legal mandate to remove environmentally damaging subsidies, and has already implemented the Global Environmental Facility (GEF) as a structure for transferring subsidy to developing countries. Since its inception in 1992, the GEF has invested over \$1 billion in biodiversity projects, most of which are related to protected areas. But although the GEF offers a mechanism for the transfer of subsidy from rich to poor countries for biodiversity conservation, the magnitude of the programme outlined here means that an enormous increase in institutional capacity is needed. Requirements include developing agencies for managing protected areas, boosting taxonomic capacity, purchasing conservation areas and building mechanisms to compensate local communities.

These institutional needs suggest that it will be at least 30 years before the changes we propose here could be fully implemented. However, given the rate at which biodiversity

is being lost, such a programme must be started now. At the international level, countries that are parties to the Convention on Biological Diversity must take the first step towards reducing environmentally damaging subsidies while committing new resources to global biodiversity conservation. The cost of global conservation is well within our means — the obstacle to progress is the lack of political will to change patterns of government expenditure.

Alexander N. James is in the Department of Land Economy, University of Cambridge, 19 Silver Street, Cambridge CB3 9EP, UK.

e-mail: anj1000@cam.ac.uk

Kevin J. Gaston is in the Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

e-mail: k.j.gaston@sheffield.ac.uk

Andrew Balmford is in the Conservation Biology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

e-mail: apb12@cam.ac.uk

- Costanza, R. et al. *Nature* **387**, 253–260 (1997).
- Myers, N. *Nature* **392**, 327–328 (1998).
- van Beers, C. P. & de Moor, A. P. G. *Addicted to Subsidies* (Institute for Research on Public Expenditure, The Hague, 1999).
- James, A. N., Green, M. J. B. & Paine, J. R. *Global Review of Protected Area Budgets and Staff* (WCMC, Cambridge, UK, 1999).
- IUCN *Parks for Life: Report of the IVth World Congress on National Parks and Protected Areas* (IUCN, Gland, 1993).
- IUCN 1997 *United Nations List of Protected Areas* (WCMC/IUCN, Cambridge/Gland, 1998).
- Balmford, A. & Gaston, K. J. *Nature* **398**, 204–205 (1999).
- Adams, J. S. & McShane, T. O. *The Myth of Wild Africa: Conservation Without Illusion* (Univ. California Press, Berkeley, 1992).
- Brandon, K. et al. (eds) *Parks in Peril: People, Politics, and Protected Areas* (Island Press, Washington DC, 1998).
- van Schaik, C. P. et al. in *Last Stand: Protected Areas and the Defense of Tropical Biodiversity* (eds Kramer, R., van Schaik, C. P. & Johnson, J.) 64–89 (Oxford Univ. Press, New York, 1997).
- United Nations Agenda 21: *Rio Declaration and Forest Principles*. Post-Rio Edition (United Nations, New York, 1993).
- Jenkins, T. N. *Future Harvests. The Economics of Farming and the Environment: Proposals for Action* (Council for the Preservation of Rural England/WWF, London/Godalming, 1990).
- Pretty, J. *What Price Wildlife in the Farmed Landscape?* (Univ. Essex, Colchester, 1999).
- FAO *FAO Production Yearbook Vol. 48*. 1994 (FAO, Rome, 1995).

Up the river without a paddle?

The theory that polio-vaccine researchers are responsible for AIDS is leaky.

The River: A Journey Back to the Source of HIV and AIDS

by Edward Hooper
Penguin: 1999. 1,070 pp. £25 (hbk)

John P. Moore

Was John F. Kennedy shot from the grassy knoll by a Mafia hit man? Did HIV-1 enter humans via a contaminated oral polio vaccine (OPV) during mass vaccination campaigns in central Africa in the late 1950s? The second of these theories, the 'OPV-HIV' hypothesis, is the subject of this massive book by Edward Hooper which has attracted considerable media attention. The BBC even issued a press release with the inflammatory title "Scientists started AIDS epidemic".

The River is, in many ways, superb. It is scholarly, thoroughly researched, well (if densely) written and deserves, indeed demands, to be taken seriously. It takes the OPV-HIV hypothesis way beyond its early manifestations in the magazine *Rolling Stone*, and infinitely past the strange Internet discussions of Californian AIDS activists. These conspiracy theorists have given the hypothesis a bad name — Hooper has redressed the balance by presenting careful and thorough research. His description of the early days of the African and Western AIDS epidemics is marvellous, but it is his support for the OPV-HIV hypothesis that will attract most attention.

The problem is that, like the grassy knoll theory, there is no smoking gun. Despite diligent investigation, Hooper cannot prove what he proposes and many of his arguments are highly speculative. Thus, a chain of events is outlined in which each link is weak. The argument goes like this. First, sick chimpanzees housed at Camp Lindi in the Congo in 1957–58 for use in medical research might have carried a primate immunodeficiency virus (PIV). Yet PIVs are not known to cause disease in chimps, and even if these animals were somehow infected with a pathogenic PIV, one could argue that any transmission to humans occurred by a more conventional route such as via blood during butchery procedures, or from biting or scratching during handling. Instead, Hooper suggests that "it could be that [kidneys from these chimps] ended up at the Wistar" — a laboratory in Philadelphia where polio vaccines were manufactured — where they contaminated vaccines with a PIV.

Chimp kidneys were shipped from Lindi to Philadelphia in 1958 and 1959. But they went to an institute doing hepatitis research, not to the Wistar, and the surviving Philadel-



Exchanging one killer disease for another?

phia- and Congo-based scientists deny that polio vaccines were ever made from chimp kidneys.

Second, the supposedly PIV-contaminated vaccine was then used in the Congo in 1957–58, transmitting the virus that evolved into HIV-1. Although PIVs can spread orally, it is unlikely that kidney-cell cultures, even if they did contain significant quantities of lymphocytes and macrophages, would produce much virus and that enough of this could survive storage and shipment.

Finally, the theory has the infected humans progressing to AIDS over the next 15–20 years, dying only around the time when the disease first became visible in central Africa in the mid-to-late 1970s. Again, this is possible, but not probable. The first authentic HIV-1 sequence, from the Congo in 1959, has characteristics making it unlikely that a chimpanzee virus, such as SIVcpz, first crossed to man only a few years earlier. A few decades is a more reasonable estimate.

One can sense Hooper's frustration that, despite all his work, he cannot prove his central point. He states "it could have happened this way", but can any of the complex sequence of events ever be proven? Hooper outlines a series of investigations, some of which are worthwhile. If additional records of polio-vaccine production exist, they should be released. And if frozen samples of the relevant vaccine stocks still exist in American or European institutions, these should be tested for the presence of possible HIV-1 precursor viruses. If the air of suspicion can be cleared, then it should be. But any such tests would have to be most carefully designed, executed and interpreted, and it is notable that some frozen vaccine stocks were analysed in Swe-

den in 1995, with negative results.

Hooper rightly argues for more research on the earliest stages of the HIV-1 epidemic in Africa and the West, with searches for archival tissues that could be analysed for HIV-1 sequences. Certainly, more knowledge of the animal reservoirs for HIV-1 and HIV-2 would be useful, and might finally settle the issue of the origin of AIDS. But the need to address the past must be balanced against the more pressing requirement to prevent HIV-1 spreading — by, ironically, the development of an effective vaccine.

Some of Hooper's proposals are less valuable. It would not, as he suggests, be informative to try to reconstruct the route of contamination. Experiments along these lines were performed in 1993 by John Garrett and colleagues in the United Kingdom. The results were negative. Even so, Hooper dismisses them as perhaps not reproducing the conditions used for polio-vaccine preparation in the 1950s. But could the conditions of 40 years ago ever be reproduced well enough?

The scientist most criticized in *The River* is Hilary Koprowski, formerly of the Wistar Institute. Many of Koprowski's actions and attitudes are presented in a very poor light. The accuracy of this portrayal can perhaps only be judged by those who know him. But even if the OPV-HIV link were correct, neither Koprowski nor anyone else would be to blame. The risks of cross-species viral transmission were much less well understood 40 years ago; one cannot condemn the past for not following the standards of the present. At worst, what Hooper argues happened would have been a tragic accident. What is it about AIDS that makes people seek a villain(s)? Do we blame anyone for the Black Death? Or for the influenza pandemic of 1919, a classic example of how animals are a reservoir for human viruses? Of course not. But in the AIDS epidemic, normal reasoning often gets discarded.

There are lessons to be learned from *The River*. One critical point is best made by a Philadelphia polio-vaccine researcher of the 1950s, who said "even though procedures [like vaccination] are very laudable and necessary, you do in fact have to make every possible effort to ensure that all safety procedures are satisfied". The theoretical and actual dangers of cross-species viral transmission are now clear, so absurdities such as baboon liver transplants into AIDS patients simply must be avoided in future. There can be few excuses today for meddling with viruses in such a potentially lethal manner.

My biggest concern over this book is that it could reinforce public distrust of science

and scientists. It is a dangerous policy to hammer science for unproven — and probably unprovable — events. Eventually the public may turn to those who believe that science is bad for society. That really would be a tragedy.

So, while respecting Hooper's scholarship and thoroughness, I am not convinced by his central argument. At most, I turn to the Scottish legal verdict of 'Not Proven'. But an outright acquittal for the polio-vaccine researchers of the 1950s would be sounder, pending the discovery of solid, and not just circumstantial, evidence to the contrary. Instead, I believe that HIV-1 and HIV-2 crossed to humans from chimpanzees and sooty mangabeys, respectively, probably when PIV-infected animals were butchered for human consumption. Oh, and I also believe that Kennedy was shot by Lee Harvey Oswald from the sixth floor of the Texas Book Repository. ■

John P. Moore is at the Aaron Diamond AIDS Research Center, Rockefeller University, 455 First Avenue, New York, New York 10016, USA.

And some big steps for science

Exploring the Moon: The Apollo Expeditions

by David M. Harland

Springer: 1999. 411 pp. £19, \$39.95

Jack O. Burns

The arrival on my desk of David Harland's *Exploring the Moon* coincided almost to the day with the thirtieth anniversary of the first manned landing on the Moon — Apollo 11: a good time to review the accomplishments of one of humankind's greatest technological achievements. Although motivated by international politics, the Apollo programme produced a surprising amount of important science about the Moon; its astronauts conducted the first field geology on an extraterrestrial body; and it greatly challenged technology and theories of the origin of the Moon that had been developed in the 1960s.

Many previous books, articles and documentaries have told the story of the exploration, astronaut training and technical hurdles during the missions. Harland uses an interesting and generally successful approach for reviewing the science from Apollo. He uses the astronauts' own words to place the reader in the shoes of these amateur and professional lunar geologists. He carefully edited scores of transcripts to produce highlights of the lunar traverses, especially the last three Apollo missions. These are interspersed with many beautiful photographs of the Moon's surface not previous-

ly seen in popular books. The result is an adventure story told by the explorers — the Magellans of the mid-twentieth century. Although the dialogue becomes tedious at times, filled with NASA jargon, it does reflect science in the making.

Each of the Apollo missions added much to our understanding of the Moon. The first large rock brought back by Apollo 11 ruled out the 'cold Moon' theory, which proposed that the maria, or dark plains, were the result of melting following meteor impacts. Instead, episodic volcanism produced the maria. Apollo 12 carried equipment for the first science experiments on the lunar surface. The lunar laser-ranging retroreflectors on this and subsequent missions established that the Moon is moving away from the Earth at a rate of 3.8 cm per year. Apollo 12 also discovered the enigmatic KREEP (abundance of potassium, phosphorus and rare-earth elements) during its mission. Apollo 14 demonstrated the need for a lunar roving vehicle for field geology. This was test-driven during Apollo 15. This mission discovered the 'genesis rock', which was later determined to be 4.1 billion years old, far older than any rocks on Earth.

Seismic experiments on this and subsequent missions recorded 'moonquakes' of magnitude 1 to 2 on the Richter scale, generated by tidal stresses in the lunar rock. These experiments also demonstrated that the Moon is a highly differentiated body, like the Earth, with a small core, a mantle and a crust. Apollo 16 results suggested that few, if any, lunar mountains were volcanically constructed, but were instead impact crater rims. It also carried the first and only telescope to the Moon, a far-ultraviolet camera that took pictures of star-formation regions in the Milky Way's nearest neighbour, the Magellanic Clouds.

The Apollo 17 crew included the first PhD geologist on the Moon, Jack Schmitt.

His insights on the lunar terrain were sophisticated and greatly furthered our understanding of the geology of the Taurus-Littrow valley. This mission demonstrated the great potential of future field geology expeditions to the Moon.

The Apollo astronauts were extraordinary men in extraordinary times. All but one were military officers and test pilots. As revealed by Harland, they struggled to find appropriate terminology and samples with geological significance, even with NASA's extensive training. Schmitt's lucid technical conversations with geologists in the 'back room' at Houston were in striking contrast to those of earlier Apollo missions. However, one marvels even today at the amazing range of tasks performed by these Apollo astronauts — piloting an advanced-technology lander using a computer far less capable than today's home PC; their expert photography of the lunar landscape and the gruelling lunar field expeditions kilometres away from the landing vehicle.

I had the great pleasure of getting to know two of these astronauts after their Apollo missions, and developed a deep respect for their courage and intelligence. As correctly described by Harland, Pete Conrad was a daredevil pilot who later commanded the remote-controlled Delta Clipper in the New Mexico desert and was helping to pioneer a new generation of single-stage-to-orbit commercial spacecraft when he recently suffered an untimely death. Jack Schmitt served a term as a US Senator for New Mexico and is today a passionate advocate of a manned expedition to Mars. In general, Harland does a good job of capturing the personalities of these men in his book.

Exploring the Moon reveals new aspects of the important scientific discoveries made by the Apollo lunar expeditions. But it leaves one thinking that we travelled to the Moon at least two decades too early and for



Rock science: the geologist Jack Schmitt performing a surface experiment during the Apollo 17 mission.

the wrong reasons. The technology and political climate of the 1960s were simply too challenging to maintain a presence on the Moon. We therefore missed out on the most important science that would have emerged from a sustained lunar base. Lunar observatories peering at the cosmos from that pristine surface, advanced field-geology expeditions and basic physics experiments searching for gravitational waves await us in the future. ■

Jack O. Burns is at the Office of Research and in the Department of Physics & Astronomy, University of Missouri, 205 Jesse Hall, Columbia, Missouri 65211, USA.

Taking the horror out of shocks

Electroshock: Restoring the Mind

by Max Fink

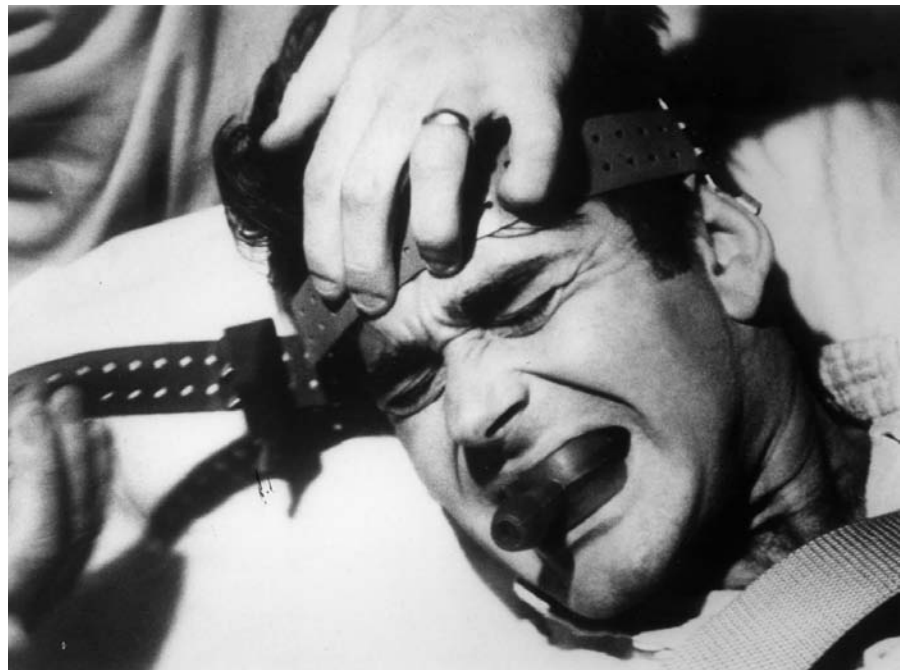
Oxford University Press: 1999. 157 pp.
£13.95, \$22

Hugh Freeman

In an America long besotted with psychotherapy, Max Fink carried an often lonely torch for the physical treatment of psychiatric disorders. Electroconvulsive therapy (ECT) has been in worldwide use for 60 years and has proved to be one of the safest procedures in medicine, with a very high success rate in appropriate cases. One might have thought that this would have given it a good public image, but no amount of research seems able to dispel the myth — largely promoted by anti-psychiatrists — that ECT damages the functioning of the brain. People mostly take their cue from alarming images of it such as those seen in the film *One Flew Over the Cuckoo's Nest*, rather than from scientific results or the testimony of millions of people who have been helped by the treatment.

Fink has written a non-technical account of ECT primarily to inform the public, but also to challenge those many mental-health professionals (particularly in the United States) whose knowledge of it is inadequate or out of date. *Electroshock* is a slim volume, packing a punch of controversy. Critics have tended to claim sarcastically that induced fits were introduced on the basis of a wrong theory, and that if they do work, no one knows how. If that is so, it certainly wouldn't be either the first or last time this has happened in medicine. The art is essentially empirical, and many interventions that have helped patients have their rationales elucidated only much later.

In fact, Fink points out that the whole thing wasn't nearly as daft as is often made out. The background was a total absence of effective treatment for mental disorders in



Electrifying image: films such as *Shock Treatment* give this useful therapy a sensationalist name.

the early twentieth century. Since Pasteur had shown that high temperatures could destroy bacteria, the induction of a new disorder (fever) might antagonize an existing one. On that basis, Julius Wagner-Jauregg in Vienna successfully treated neurosyphilis with malaria. It had also been observed that schizophrenic patients occasionally recovered after developing epilepsy through a head injury or infection. Ladislaus Meduna in Budapest therefore proposed that artificial epilepsy might be a useful treatment for schizophrenia. This was long before any antipsychotic drugs were available.

As is now well known, the results in the treatment of schizophrenia were mixed, but when the same procedure was tried in patients with severe depression, its effect seemed miraculous — some 20 years before the first antidepressants. In time, the original convulsive injections were replaced by a more acceptable electrical current, while anaesthesia, muscle relaxants and continuous oxygen removed most of the unpleasant aspects of the procedure. In folk memory, though, and in many media representations of ECT, things have hardly changed.

One major drawback of Fink's account is an overly American focus. Fink describes how ECT tended to be abandoned with the rise of psychotherapy and psychopharmacology, was "rediscovered" in the 1970s, is still ignored in university teaching centres and is still legally and financially restricted.

But in Europe, practically none of this happened, although local restrictions occurred in The Netherlands and similar regulations have just been introduced in Italy, and restrictive regulations have just been introduced in Italy. Where Britain failed, as Fink rightly points out, is in allow-

ing obsolete or poorly maintained equipment to remain in use, not providing electro-encephalographic monitoring, and not training young psychiatrists in the method adequately. This reflects the lack of resources afflicting the UK mental-health services. In the United States, things are now even worse, with few public or Veterans hospitals offering ECT at all and the rigid treatment algorithms of 'managed care' having little place for it.

Professionals will find the clinical and scientific parts of Fink's work more controversial. For a start, he believes he knows how it works. In mental illness, "the hormones are wildly disordered", but impulses from the central part of the brain in an induced fit stimulate the hypothalamus to discharge products that gradually regulate the whole endocrine pattern of the body. It may or may not be so, but proof is needed; and the immensely complex actions of the neurotransmitters seem to have been ignored.

Fink advocates ECT for a wider range of disorders and in much longer courses than is advised in other standard references. However, his choice of clinical response as the essential guide would be generally accepted. It is true that ECT often doesn't get a fair trial because clinicians give it up too quickly. But not much attention is given to an experimental method that may well replace ECT. Rapid transcranial stimulation aims to produce the same effects without the need for a fit or unconsciousness. It's too soon to say whether this will mean the end of ECT, but until then, Fink is right to try and persuade us to make the best use of what we have. ■

Hugh Freeman, former editor of the *British Journal of Psychiatry*, is at Green College, Oxford OX2 6HG, UK.

Border control at the frontiers of science

Cultural Boundaries of Science: Credibility on the Line

by Thomas F. Gieryn
University of Chicago Press: 1999. 398 pp.
\$21, £16.75 (pbk)

N. David Mermin

"Science is the systematic enterprise of gathering knowledge about the world and organizing and condensing that knowledge into testable laws and theories. The success and credibility of science is anchored in the willingness of scientists to: (1) expose their ideas and results to independent testing and replication by other scientists; this requires the complete and open exchange of data, procedures, and materials; (2) abandon or modify accepted conclusions when confronted with more complete or reliable experimental evidence. Adherence to these principles provides a mechanism for self-correction that is the foundation of the credibility of science."

"When one arrives at the mathematical theories on which quantum mechanics is based, one realizes that the attitude of certain physicists in the handling of these theories truly verges on delirium. ... One has to ask oneself what remains in the mind of a student who has absorbed this unbelievable accumulation of nonsense, real hogwash! It would appear that today's physicists are only at ease in the vague, the obscure, and the contradictory."

Thomas Gieryn tells us that specimens such as those above are exercises in "cultural cartography", or "boundary work". The first was hammered out by the Panel on Public Affairs (PPA) of the American Physical Society to help the public distinguish pseudo-science from the real thing. The second was advanced by the mathematician Jean Dieudonné, deploring what physicists try to pass off as mathematics. Both the PPA and Dieudonné are drawing borders around acceptable practice. Physicists find themselves within the borders of the PPA's definition, but outside Dieudonné's.

Those engaging in such boundary work may believe that they are trying to distinguish between effective and ineffective ways of acquiring or formulating reliable knowledge. But Gieryn considers that it is not knowledge that is at stake, but epistemic authority — recognition that one is a rightful purveyor of knowledge. If scientists see it differently, this is not because of self-interested deceit: "[M]ore often than not, I suspect, scientists really believe that their representations of science tell it like it is. ... Many believe that the epistemic authority of science is justified ... by the unique, necessary, and universal

elements of its practice — behaviors, dispositions, methods, rules, tools, and languages that simply work best to make truth. ... [But] credibility in the culturescape is not decided in tinkering at the lab bench or in the refereeing of a manuscript or in the machinations of instruments, statistics, or logic."

Gieryn offers five case studies of credibility seeking.

● John Tyndall's "double boundary work" in trying to distinguish nineteenth-century science from both religion and engineering. The practical utility of science counted for much in demarcating the first border but little in establishing the second, while science as a flowering of human culture made little appearance on the first border but was much in evidence on the second.

● Whether social science belongs in the same territory as natural science depended on whether one was debating its proposed inclusion in the new US National Science Foundation shortly after the Second World War, or arguing 20 years later about the establishment of a separate National Social Science Foundation.

● The debate in 1836 over whether the chair of logic and metaphysics at the University of Edinburgh should be awarded to an eminent phrenologist or a traditional logician. "In what world of meanings did ... phrenology seem plausible, truthful, useful, scientific, and chair-worthy?"

● The very public 1989 disputation over cold fusion, in which one of the tests of proper science was whether members of the press were or were not granted priority over scientists during post-presentation question periods.

● A long examination of what Gieryn very much enjoys calling the science of "imperial economic botany", which starts in a Cambridge laboratory, and concludes, via the West Indies and after a quarter of a century in India, back in the England of the Second World War, as a crusade for the use of compost over that of artificial fertilizers.

In an epilogue, Gieryn views the recent 'science wars' between scientists and those who study science as more boundary work. (Here I have to report that he errs in taking a remark I once made about the nature of science teaching to be about the nature of science.) It is tempting for a physicist, reviewing a book by an acknowledged social constructivist, to view the book across that particular border, and since Gieryn himself brings the matter up, there is no need to resist.

Am I, then, appalled that Gieryn has, through his scientific studies of science, "come to see such concepts [as 'rational', 'empirical', 'modern' and 'science'] not as a set of rules for proper fact-construction, but as rhetorical tools deployed in the pursuit or defense of epistemic authority, or in efforts to deny legitimacy to rival claims"? No, I am not. He makes a compelling case that the concepts are repeatedly used by scientists

in precisely such ways. I do not find this as interesting as he does, but then, why should I?

One of my personal idiosyncracies is that I can't help wanting to know if dispositions really are so localized in the brain that, when highly developed, they can raise bumps on the skull, if deuterium really did undergo fusion in those palladium electrodes, whether children really are healthier if they're raised on organically fertilized food. Questions like these are outside the boundaries of Gieryn's kind of social science. It's hard to tell whether he finds such extraterritorial matters also boring, but they are clearly irrelevant to what interests him professionally.

So, for me, Gieryn's clinical detachment makes him a less interesting storyteller. But I would say the same about the stultifying convention that banishes any hint of human activity from most research papers in fields closer to my own. Indeed, if one outcome of the science wars were to make physicists less uncomfortable with using rhetoric when describing their work, and sociologists less queasy about objective facts, life could become a great deal more entertaining on both sides of the border. ■

N. David Mermin is in the Department of Physics, Cornell University, Ithaca, New York 14853-2501, USA.

New in paperback

How Nature Works: The Science of Self-Organized Criticality

by Per Bak
Springer, \$18, £15.99

"This book is essential reading for those interested in complex systems in general. Its idiosyncratic personal style may intrigue as many as it turns off, but, like the similarly flawed classic book of Mandelbrot, it will reward a sufficiently sceptical reader." Philip W. Anderson, *Nature* **383**, 772–773 (1996)

Isaac Newton: The Last Sorcerer

by Michael White
Perseus Books, \$16

"Although Newton is today chiefly remembered for his theoretical work in *Principia*, he also spent much of his time doing experiments, even using equipment — or so argues Michael White in *Isaac Newton: The Last Sorcerer*." *Nature* **392**, 669 (1998)

Endless Frontier: Vannevar Bush, Engineer of the American Century

by G. Pascal Zachary
MIT Press, \$22, £15.50

"G. Pascal Zachary has written the first definitive and scholarly biography of Bush (1890–1974) ... The work is extraordinarily well documented, with 73 pages of notes and a seven-page list of books and other sources." Harvey Brooks, *Nature* **390**, 34–35 (1997)

From horsehair to lightning rods

Analogy is a powerful tool in Benjamin Franklin's natural philosophy.

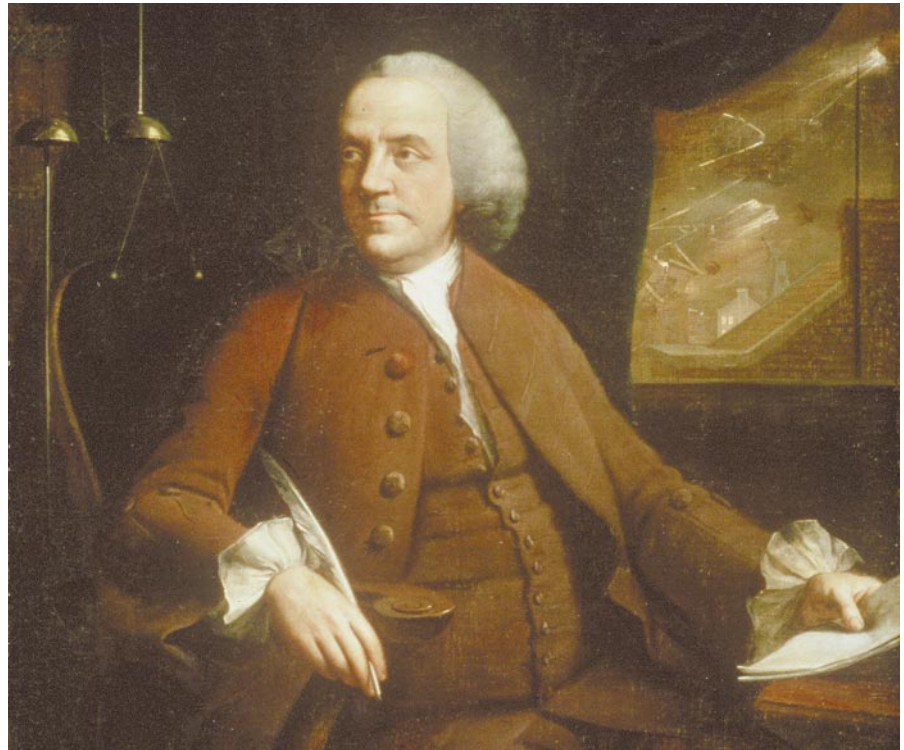
J. L. Heilbron

Anal^Aogical reasoning and the concept of conservation are among the most powerful tools of science. They first took on quantitative form in physics in the eighteenth century. Among the useful analogies developed then were parallels between the laws of gravity and electrostatics and between trajectories of particles and rays of light. Meanwhile, philosophers unscrambled various old notions of 'force' into the conservation of *vis viva* (kinetic energy) and momentum — and a retired printer at the edge of the civilized world used analogy to discover a brand new conservation law.

The printer was Benjamin Franklin. He began with a dour principle, itself a conservation law in moral philosophy: the sum of pain and pleasure in this world is always exactly zero. The truth of the principle may be more obvious than its bearing on science. An example given by the propounder of the principle provides a clue. Let a man have ten degrees of pleasure: ten degrees of pain are therefore debited to his account. Let him pay up: he returns to his original indifferent state, in which a non-sentient being would have remained all along. Is it not obvious that pleasure corresponds to positive electricity, pain to negative and dullness to the neutral state? The conservation of pain and pleasure appeared in Franklin's writings more than 20 years before the conservation of electrical charge.

Similarly, although a philanthropist by instinct, Franklin reasoned that nothing could be done to improve the situation of the poor, since they already received all the income from the land: all the income that, on physiocratic principles, exists. The wealth of the nation consists of two parts, he reasoned: savings and "clear revenues". Planting and mining create a positive balance against which the maintenance of labourers must be debited — according to Franklin's guiding book-keeping analogy — in equal amounts. Accumulated surplus will not help the poor either, since it cannot add to wealth unless invested in the production of raw material. One man buys land and becomes positively charged, agriculturally speaking, simultaneously suffering a deficit in his wallet; the seller goes negative in land and positive in cash, all amounts exchanged in exact equality. The working of new land brings nothing new; the poor may be more numerous but they are no richer.

Franklin's difficulty in conceiving of accumulations without compensatory



Franklin: 'gripped by the mode of thought that created the concept of electricity plus and minus'.

deficits indicates how strongly his mind was gripped by the mode of thought that created the concept of electricity plus and minus.

Many of Franklin's other contributions to natural philosophy also derived from analogies. The best known of these contributions is his conjecture of the identity of lightning and ordinary electricity. He based his case on 12 counts of analogy: colour of light, conduction by metals, rending of bodies, firing of inflammable substances, and so on. Thence came his proposal to preserve buildings from damage in a thunderstorm by outfitting them with pointed metal rods. To explain what he called the "power of points" — their aptness in "drawing off and throwing off the electrical fire" — he appealed, naturally, to analogy.

"As in plucking the hairs from a horse's tail, a degree of strength not sufficient to pull away a handful at once could yet easily strip it hair-by-hair; so a blunt body presented cannot draw off a number of particles [of electrical fire] at once, but a pointed one, with no greater force, takes them away easily, particle by particle."

Those who found this reasoning persuasive mounted a pointed metal pole above their houses and ran its bottom into moist ground, believing that it would silently steal the electricity of every passing thunder cloud. Some, who could not follow the anal-

ogy between stripping horse's tails and despoiling thunder clouds, concluded that lightning rods courted the very strikes they were intended to deflect. However well they reasoned, Franklin's intuitive, far-fetched analogizing reached deeper. Lightning rods work — though not by the power of analogy.

Elsewhere in his natural philosophy, Franklin rose from the mechanics of a stove to the causes of the winds, the origin of the Gulf Stream, the basis of water spouts, and a way to preserve meat during the summer. This contribution to refrigeration emerged from a concatenation of analogies. Franklin knew that a thermometer could be made to fall by wetting its bulb and directing a stream of air at it. "It seems by this, that a Man naked, and standing in the Wind, and repeatedly wet with Spirits, might be frozen to death in a Summer's Day." On the same principle, wrap your meat in a damp cloth and hang it in the chimney, where, according to Franklin, who had a chimney like mine, a slight breeze is nearly always blowing.

In 1771, the *Encyclopaedia Britannica* declared: "a great part of our [natural] philosophy has no other foundation than analogy". That was true of Franklin's physics. Is it not also true of ours? □

J. L. Heilbron is at Worcester College, University of Oxford, Oxford OX1 2HB, UK.

The devil is in the distance

Bohdan Paczynski

An apparent discrepancy between two different distance measurements to a faraway galaxy may lead to a revised value for the age of the Universe.

The age of the Universe is inversely proportional to the Hubble constant — the rate at which the Universe is currently expanding. The Hubble constant, H_0 , is defined by the relation $V = H_0 d$, where V is the velocity of a galaxy at a distance d from us. So establishing a value for H_0 requires working out accurate velocities (which is relatively easy) and exact distances (much more problematic) of extragalactic objects. The difficulty in measuring such distances is highlighted by Maoz *et al.*¹ on page 351 of this issue.

Over the past few years estimates of H_0 have been in the range $60\text{--}75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (where $1 \text{ Mpc} = 3.1 \times 10^{24} \text{ cm}$), and the corresponding time the Universe has been expanding was thought to be between 13 and 16 billion years. The uncertainty in the value of H_0 , and therefore the age of the Universe, is mainly caused by errors in measuring large distances. Most techniques give only relative distances, whereas for H_0 we want absolute distances. The most reliable distance measurements are the simplest; that is, purely geometrical. In our own galaxy, distances to nearby stars are obtained from trigonometric parallax — the apparent change in position due to a changing viewpoint — but this doesn't work for stars in other galaxies.

Galaxy NGC4258 may turn out to be a better yardstick (Fig. 1). Its nucleus was found to contain strong water maser sources (the microwave equivalent of lasers)², which form a disk rotating around a compact massive object, presumably a black hole^{3,4}. Last month in *Nature*, Herrnstein and co-workers⁵ reported measurements of the orbital motion of these water maser sources, from which they calculated two somewhat independent distances to galaxy NGC4258. The first was based on the acceleration of the masers, and the second on their proper motion (their apparent angular velocity), which, thanks to the simple geometry of the disk, can be converted into extragalactic distances. The two methods gave almost identical distances to the nucleus: 7.1 Mpc and 7.2 Mpc. With this geometric technique, the distance to a faraway galaxy has been measured more accurately, and almost certainly more reliably, than the distance to any other extragalactic object.

Maoz *et al.*¹, however, report a different distance measurement to NGC4258. Using the Hubble Space Telescope (HST) Maoz *et al.* have discovered 15 giant pulsating

stars, known as Cepheids, in this galaxy and recorded variations in their brightness. Cepheids make useful 'standard candles' because they pulsate in a way that depends precisely on their true brightness, which means that their apparent brightness viewed from Earth indicates their distance from us. Over 800 of these standard candles have been measured as part of the HST Key Project, which set out in 1991 to redefine H_0 based on Cepheid distances to 18 galaxies. A few months ago they announced their most definitive value yet for H_0 : $68 \text{ km s}^{-1} \text{ Mpc}^{-1}$, with an estimated uncertainty of 10%.

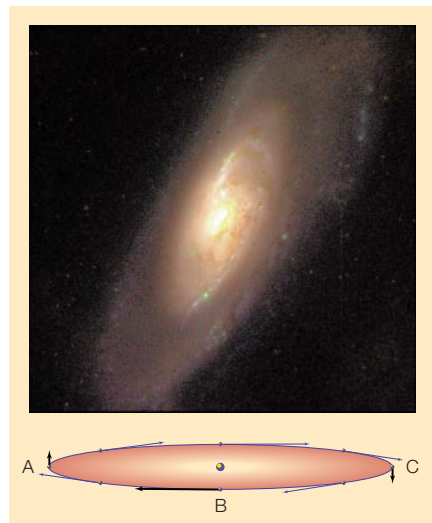


Figure 1 A better cosmic yardstick? A geometric distance to galaxy NGC4258 reported by Herrnstein *et al.*⁵, combined with the discovery of Cepheid variable stars in this galaxy by Maoz *et al.*¹, may lead to a revised value for the age of the Universe. The geometric distance is based on the motion of water maser sources in the nucleus of NGC4258. The maser disk is made of a large number of concentric rings with varying inclination and orientation. The geometry of one of the rings with an inclination angle of 82° (almost edge on) is shown here. The ring is assumed to rotate because of gravitational attraction by the central mass, presumably a super-massive black hole. The strong maser sources are observed at locations A, B and C, as shown with black arrows. Those at A move away from us with a velocity close to $1,000 \text{ km s}^{-1}$; those at C come towards us with a similar velocity; and those at B are expected to move transversely. Locations A and C are the best for measuring radial velocity, whereas B is the best for measuring acceleration and proper motion.

But when Maoz *et al.* apply the HST Key Project methods to the Cepheids in NGC4258, they find the distance to that galaxy is $8.1 \pm 0.4 \text{ Mpc}$, which differs by about 12% from the maser distance, despite reported errors in both distances of between 5% and 7%. The authors conclude that if you accept the maser distance as the more accurate one, then the value of H_0 obtained by the HST Key Project will have to be increased by 12%, reducing the age of the Universe (all other cosmological parameters being equal) by 12% to about 12 billion years.

Things may not be as bad as they look. The Cepheid distance scale used by the HST Key Project relies on the assumption that a neighbouring galaxy, the Large Magellanic Cloud (LMC), is 50 kpc from us. The distance to the LMC is the benchmark by which all other extragalactic distances are calibrated. Although 50 kpc was the value adopted by the HST Key Project, the most direct methods available today give a much shorter distance to the LMC. The direct methods are based on observations of different types of stars that are abundant in the LMC: eclipsing binaries^{6–8}, pulsating stars of the RR Lyrae type^{9,10} and the so-called red-clump giants^{11–13}, and they all give the same answer: the LMC is $44.5 \pm 1.0 \text{ kpc}$ away. This 'short distance' can be used to calibrate the brightness of Cepheids in the HST Key Project¹⁴, which then agrees with the brightness of Cepheids observed in NGC4258, given the accurate maser distance obtained by Herrnstein *et al.*⁵. The assumption that the LMC is 44.5 kpc away increases the Cepheid-based value of H_0 by 12% to about $76 \text{ km s}^{-1} \text{ Mpc}^{-1}$, bringing it in line with the maser-based estimate.

This is not the end of the story. A recent analysis of Cepheids in the nearby giant Andromeda galaxy (M31) revealed that most of their images are blended with images of close companion stars, which makes Cepheids appear brighter than they really are¹⁵. The farther away a galaxy, the more difficult it is to recognize that the stellar image is blended. It is not known how large the blending effect is for the HST Cepheid images in distant galaxies, but whatever the size of the effect the result is clear: the true apparent luminosities must be smaller than measured, and so the galaxies must be farther away than thought. This has the result of reducing H_0 , so the two effects — overestimating the LMC distance and neglecting blending — will partly compensate for each other. It is even possible that

these effects almost cancel, and we may not need to revise our value of H_0 after all.

My assessment of the various systematic effects is a personal one, and not everyone agrees with it. For instance, according to a recent review article¹⁶ the distance to the LMC is 55 kpc, which is even larger than the value adopted by the HST Key Project. Clearly, systematic effects and personal judgements dominate published values of H_0 . This may get better following the release of 340,000 photometric measurements of 1,300 LMC Cepheids by the OGLE (Optical Gravitational Lensing Experiment) team¹⁷, which are available for everyone to analyse using their favoured methodology. ■

Bohdan Paczynski is at Princeton University

Observatory, 124 Peyton Hall, Princeton University, Princeton, New Jersey 08544-1001, USA.

e-mail: bp@astro.princeton.edu

1. Maoz, E. *et al. Nature* **401**, 351–354 (1999).
2. Claussen, M. J. *et al. Nature* **310**, 298–300 (1984).
3. Miyoshi, M. *et al. Nature* **373**, 127–129 (1995).
4. Maoz, E. *Astrophys. J.* **447**, L91–L94 (1995).
5. Herrstein, J. R. *et al. Nature* **400**, 539–541 (1999).
6. Bell, S. A. *et al. Mon. Not. R. Astron. Soc.* **265**, 1047–1052 (1993).
7. Guinan, E. F. *et al. Astrophys. J.* **509**, L21–L24 (1998).
8. Udalski, A. *et al. Astrophys. J.* **509**, L25–L28 (1998).
9. Gould, A. & Popowski, P. *Astrophys. J.* **508**, 844–853 (1998).
10. Udalski, A. *Acta Astron.* **48**, 113–145 (1998).
11. Udalski, A. *et al. Acta Astron.* **48**, 1–17 (1998).
12. Stanek, K. Z. *et al. Astrophys. J.* **500**, L141–L144 (1998).
13. Udalski, A. *Acta Astron.* **48**, 383–404 (1998).
14. Gibson, B. K. *et al. http://xxx.lanl.gov/abs/astro-ph/9908149*
15. Mochejska, B. J. *et al. http://xxx.lanl.gov/abs/astro-ph/9908293*
16. Feast, M. *Publ. Astron. Soc. Pacif.* **111**, 775–793 (1999).
17. Udalski, A. *et al. http://xxx.lanl.gov/abs/astro-ph/9908317*

Biophysics

Fungus punches its way in

Nicholas P. Money

The development of a specialized fungal cell known as the appressorium is a wondrous story. This microscopic structure inflates on the surface of a grass leaf, then generates enough force to push through the fatty plant cuticle and glassy epidermal cell wall to tap the juices within. Clemens Bechinger and colleagues have made a breakthrough in understanding how this infection platform works, by using an optical method to measure the forces produced by a single appressorium. This work, reported in *Science*¹, verifies earlier predictions about the magnitude of the invasive force, pushing us closer to a comprehensive picture of the penetration mechanism.

Our understanding of appressorial development draws on experiments using two of the most important cereal pathogens — *Magnaporthe grisea* (the rice blast fungus) and *Colletotrichum graminicola*. Infection is initiated when a fungal spore lands, and germinates, on the surface of the host. The appressorium develops as a swelling at the tip of the developing fungal filament (hypha), and the structure becomes bonded to the host by a strong glue^{2,3}. Maturation involves the deposition of melanin within the appressorial wall, and, as the appressorium blackens, accumulation of osmolytes such as glycerol in the cytoplasm⁴. These osmolytes draw water into the appressorium, generating extraordinary levels of intracellular turgor pressure⁵. Following this increase in turgor, a thin penetration hypha extends from the base of the appressorium and perforates the host surface.

Although we know much about how appressoria function, the absence of methods to measure the invasive forces that they exert has long posed a problem for those

studying mechanical interactions between pathogenic fungi and their hosts. Bechinger and colleagues' solution is both ingenious and elegant.

Appressoria will form not only on plant leaves, but also on a variety of hydrophobic surfaces. Some years ago, we reported⁶ that appressoria can penetrate Mylar and Kevlar (poly(ethylene terephthalate) and *p*-phenylene terephthalamide, respectively), spawning visions of bulletproof vests destroyed by airborne moulds. In Bechinger and colleagues' experiments, appressoria of *C. graminicola* formed on an aluminium surface that was part of a metal sandwich (called a waveguide) enclosing a 1- μ m-thick layer of the viscous silicone compound used in breast implants. As the fungus began its futile attempt at penetration, each appressorium deformed the aluminium and underlying gel.

The depth of these indentations (which is proportional to the applied force) was measured by illuminating the sandwich with a laser. Laser light is propagated through the waveguide, and it reflects at the upper interface between the silicone and the aluminium. So, the angle of incidence at this interface depends on any deformation of the otherwise flat aluminium. By collecting data on the intensity of the reflected light with a CCD camera, the authors built a three-dimensional image of the waveguide's surface as the laser scanned the interface. This provided them with vertical spatial resolution in the nanometre range. To determine the forces exerted by appressoria from these spatial measurements, the authors calibrated the waveguide by probing with a glass capillary of known spring constant using forces up to 35 μ N.

The appressorium derives its invasive force from its turgor pressure. This has been



100 YEARS AGO

Prof. F. P. Venable briefly discussed the nature of the elements in his address to the Section of Chemistry. He passed in review some of the evidence which leads to the belief that the so-called elementary atoms are but compounds of an intimate peculiar nature, the dissociation of which has not yet been accomplished. Referring to the conclusions to which investigations lead, it was remarked that the hypothesis that the elements are built up of two or more common constituents has a larger number of supporters, and would seem more plausible than Graham's hypothesis. Some have supposed one such primal element by the condensation or polymerisation of which the others were formed. Others have adopted the supposition of two elements. There are many practical difficulties in the way of these suppositions; the lack of uniformity in the differences between the atomic weights, the sudden change of electro-chemical character, and the impossibility, so far, of discovering any law underlying the gradation in the properties of the elements with the increase of atomic weights, are some of the difficulties. ... It is possible to think of valency as dependent upon the character of the motion of the atom, but one cannot well conceive of a similar dependence of atomic weight and all the other properties.

From *Nature* 21 September 1899.

50 YEARS AGO

In his presidential address to Section D (Zoology) of the British Association at Newcastle recently, Prof. A. C. Hardy put forward the tentative suggestion that one factor in moulding patterns of behaviour among the members of a given species might be something akin to what is usually called telepathy, a phenomenon which he considered had now been established and which was clearly a revolutionary discovery from whatever point of view it might be regarded. Although, perhaps, a little apologetic in advancing this idea, Prof. Hardy did not conceal his belief that few could reject the evidence for telepathy were they to examine it with unbiased minds and without undue prejudice. ... The study of such matters has, as he rightly pointed out, become a highly specialized and difficult pursuit which the ordinary scientific man cannot hope to follow while at the same time pursuing his own interests and investigations.

From *Nature* 24 September 1949.

measured by incubating appressoria in solutions of increasing osmolality until they deflate to atmospheric pressure, and also by measuring the melting point of intracellular ice^{5,6}. Estimates have exceeded 8 MPa or 81 atmospheres — the highest turgor pressures on record. The actual force does not depend entirely on turgor pressure, though. It is affected by both the size of the penetration hypha and the degree to which its cell wall yields: the looser the wall, the more of the force within the appressorium will be exerted on the substrate underneath⁷. According to Bechinger and colleagues, appressoria of *C. graminicola* exert an average force of 17 μN . This figure is remarkably close to estimates published for *M. grisea* (up to 8 μN) based on the pressure data⁸.

Some details of the mechanical interaction between the appressorium and resistant surfaces remain in question. Measurements of impressions made by a metal probe at various loads (the Vicker's indentation technique) have shown⁵ that appressoria can penetrate Mylar films with hardness values exceeding 200 MPa (or $2 \times 10^8 \text{ N m}^{-2}$). To relate this measurement to the process of penetration, it needs to be scaled down to the size of a penetration hypha (about 1 μm^2). The prediction from these measurements is that a cell of this size must exert 200 μN to penetrate the Mylar film. Yet Bechinger and colleagues' waveguide data indicate that the force is ten-fold lower.

How can we explain this paradox? Perhaps Vicker's method overestimates the force necessary to penetrate Mylar on a microscopic scale. Or maybe the fungus has some mechanism for softening this compound, although no enzyme has yet been identified that can degrade it. Moreover, although it can be pierced by appressoria, Mylar has proved incredibly resistant to microbial degradation in long-term trials. But even if appressoria do not operate by mechanics alone, 17 μN is a tremendous force for a cell — if a force of 17 $\mu\text{N} \mu\text{m}^{-2}$ were exerted over the palm of one hand, a human could lift an 8,000-kg school bus. Bechinger and colleagues' report has, then, added to a picture of awesome microbial power that was already quite familiar to fungal biologists. ■

Nicholas P. Money is in the Department of Botany, Miami University, Oxford, Ohio 45056, USA.

e-mail: moneynp@muohio.edu

1. Bechinger, C. *et al.* *Science* **285**, 1896–1899 (1999).
2. Howard, R. J. in *The Mycota: Plant Relationships* (eds Carroll, G. & Tudzynski, P.) 43–60 (Springer, Berlin, 1997).
3. Ebata, Y., Yamamoto, H. & Uchiyama, T. *Biosci. Biotechnol. Biochem.* **62**, 672–674 (1998).
4. De Jong, J. C. *et al.* *Nature* **389**, 244–245 (1998).
5. Howard, R. J., Ferrari, M. A., Roach, D. H. & Money, N. P. *Proc. Natl Acad. Sci. USA* **88**, 11281–11284 (1991).
6. Money, N. P. & Howard, R. J. *Fungal Genet. Biol.* **20**, 217–227 (1996).
7. Money, N. P. in *Molecular Genetics of Host-Specific Toxins in Plant Disease* (eds Kohmoto, K. & Yoder, O. C.) 261–271 (Kluwer, Dordrecht, 1998).
8. Money, N. P. *Can. J. Bot.* **73**, S96–S102 (1995).

High-temperature superconductivity

Electrons pair themselves

Joe Orenstein

Soon the discovery of high-temperature superconductivity in copper-oxide materials will belong to a previous millennium. In the early days an explanation of this phenomenon was expected well before the next epoch, but the 1990s saw the mystery settle like an indigestible dinner at a roadside restaurant. Some new insights, such as the paper from Carbotte *et al.*¹ on page 354 of this issue, may offer a dose of much needed relief.

Bardeen, Cooper and Schrieffer (BCS) explained superconductivity seen in conventional metals when cooled below the transition temperature, T_c , back in 1957. The essential ingredient of their theory is the interaction between freely moving electrons and the lattice of ions that form the structural basis of the solid. This interaction leads to a net attraction and pairing of electrons. Although each electron is a type of particle known as a fermion, a pair of electrons becomes a composite boson (a particle that obeys completely different quantum statistics). Superconductivity can be viewed as a condensed phase analogous

to a Bose–Einstein condensate of atoms.

Immediately after the discovery of high- T_c superconductivity, it was expected that a reasonable theory for the copper oxides might come from 'BCS on steroids' — a beefed-up version of the same electron-lattice mechanism. This idea received support from measurements of the magnetic-flux quantum in copper oxides, which showed the superconducting condensate to be composed of electron pairs². Later advances in materials fabrication led to the dramatic discovery that the electron pairs have *d*-wave orbital symmetry rather than *s*-wave, an observation that does not fit into the electron-lattice theory^{3,4}.

With the lattice seemingly out of the picture, the search began for another type of 'glue' to bond electrons into pairs. At present no external bonding agent has been found. Without such an agent we must conclude that electrons can pair and condense despite the repulsion of their negative charges, and that they do so at spectacularly high temperatures. Can electrons act as both gluer and glue?

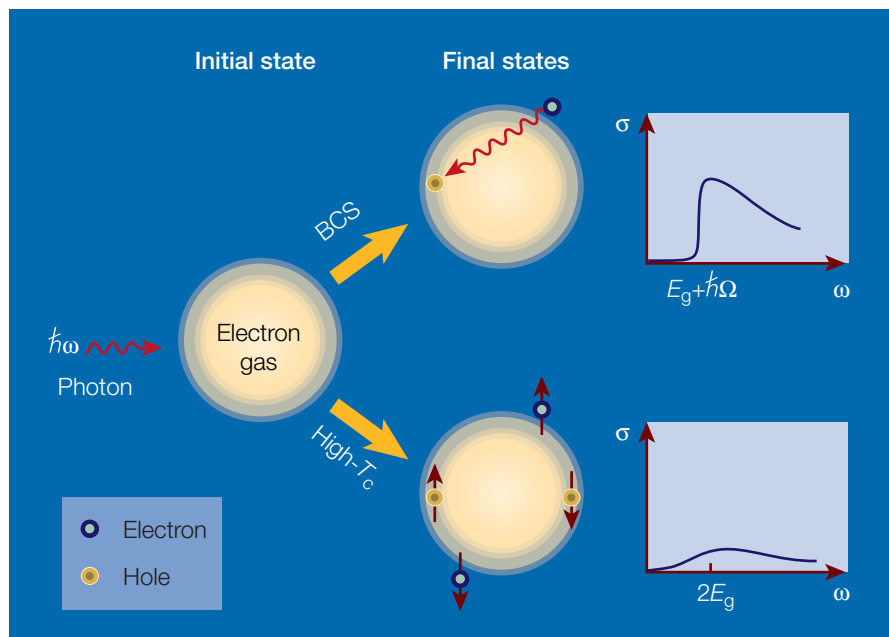


Figure 1 Absorption of a photon by a superconductor. Shaded circles illustrate the region of momentum space occupied by electrons in their superconducting ground state. Photon absorption promotes an electron to an unoccupied state, leaving a hole behind. Because the photon has nearly zero momentum, photogeneration of the electron-hole pair alone is forbidden by momentum conservation. In BCS superconductors a phonon with momentum opposite to that of the electron-hole pair is generated as well. Absorption occurs when the photon energy equals the sum of the phonon energy $\hbar\Omega$ and the gap energy E_g . In high- T_c superconductors the coupling of electrons to each other may play a more important role than the electron-lattice coupling that drives superconductivity in BCS superconductors. In this case, photon absorption generates an exotic final state comprising two electron-hole pairs.

Carbotte *et al.*¹ have found an important hint for solving this riddle in their analysis of the optical properties of the copper oxides. The optical conductivity, $\sigma(\omega)$, measures the rate at which a metal absorbs photons of energy $\hbar\omega$. Figure 1 shows the absorption process: direct conversion of a photon to an electron-hole pair is forbidden because it violates momentum conservation. But if the electrons are strongly coupled to lattice vibrations, then photon conversion to an electron-hole pair plus an energy quantum of lattice vibration (phonon) is possible. Above T_c the onset of absorption occurs at the vibrational frequency, Ω , whereas in the superconducting state it shifts to $E_g + \hbar\Omega$, where E_g is the energy required to break apart an electron pair — the so-called superconducting energy gap. In normal BCS superconductivity, where the lattice provides the pairing glue, features in the optical spectrum are a Rosetta Stone for superconductivity. Careful analysis of $\sigma(\omega)$ as a function of lattice frequency can determine the size of the superconducting energy gap, identify which lattice vibrations lead to pairing and measure the strength of the electron-lattice coupling.

So far so good, but deciphering the optical properties of high- T_c superconductors has turned out to be less simple. Optical measurements of copper oxides in the non-superconducting state have revealed few secrets, as $\sigma(\omega)$ was surprisingly featureless. But a sharp absorption onset, at an energy of about 50 meV, does appear upon cooling below T_c . The origin of this spectral feature has been a long-standing puzzle. Because its energy is similar to values of E_g determined by other techniques, it is tempting to identify this absorption onset with the direct photogeneration of an electron-hole pair. But momentum conservation dictates that the energy threshold for this process must be E_g plus the energy of the 'gluon' which binds the pair.

Soon after these measurements, it was suggested that, in the absence of a lattice excitation, the absorption threshold for high- T_c superconductors occurs at $2E_g$ rather than E_g (ref. 5). This is because the lowest-energy final state with zero momentum must contain two electron-hole pairs, each pair making an equal and opposite contribution to the total momentum. Even so, careful calculations predicted a soft rather than a sudden onset of absorption because zero-momentum states of four particles should have a broad range of energies⁶.

Carbotte *et al.*¹ have seized on evidence from inelastic neutron-scattering measurements to reopen the case. Such experiments have revealed the existence of an excited state 41 meV above the superconducting ground state. The excitation consists of an electron-hole (or electron-electron) pair in a spin-aligned or 'triplet' state⁷⁻⁹. Rather than forming a continuum as expected, the energy of

these pairs is concentrated in a remarkably narrow range centred on 41 meV. As a result, optical excitation of a four-particle state may produce a sharper onset than previously thought.

To analyse the optical spectrum, Carbotte *et al.* make a bold leap: they treat the 41 meV resonance exactly as a phonon in the BCS theory of superconductivity. Properly analysed, BCS theory instructs you to focus on the second derivative of the conductivity with respect to frequency. Plotting this should give a faithful representation of the spectrum of excitations responsible for pairing, but shifted in energy by half the energy gap. When Carbotte *et al.* overlaid the second derivative of σ (which peaks near 68 meV) and the neutron scattering spectrum shifted by $E_g/2$ (or 27 meV), they found a near-perfect match. With this correspondence in hand, they then deduced the strength of coupling of electrons to the 41 meV excitation. They conclude that the coupling is strong enough for the triplet pair to act as the long-sought binding agent.

These observations are certain to focus attention on whether the triplet resonance can cause electrons to pair in the same way as lattice vibrations do in conventional superconductors. Unlike a phonon, the 41 meV resonance is an excitation of the electronic

system itself. Using a different set of theoretical tools¹⁰, Demler and Zhang¹¹ have also suggested that the 41 meV resonance might be the culprit behind electron pairing in high- T_c superconductivity. In contrast with Carbotte *et al.* they emphasized a difference between phonons and the triplet resonance: whereas lattice vibrations are well-defined resonances above T_c , the triplet 41 meV peak appears only in the superconducting state. Despite their differences, these papers confront superconductivity researchers with the same basic question — does the 41 meV resonance drive high- T_c , or is it merely along for the ride? ■

Joe Orenstein is in the Department of Physics, University of California and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

e-mail: joeo@ux5.lbl.gov

1. Carbotte, J., Schachinger, E. & Basov, D. N. *Nature* **401**, 354–356 (1999).
2. Gough, C. E. *et al.* *Nature* **326**, 855–857 (1987).
3. Wollman, D. A. *et al.* *Phys. Rev. Lett.* **71**, 2134–2137 (1993).
4. Tsuei, C. C. *et al.* *Phys. Rev. Lett.* **73**, 593–596 (1994).
5. Orenstein, J. *et al.* in *Electronic Properties of High- T_c Superconductors* (ed. Kuzmany, H.) 254–259 (Springer, Berlin, 1990).
6. Quinlan, S. M. *et al.* *Phys. Rev. B* **53**, 8575–8582 (1996).
7. Rossat-Mignod, J. *et al.* *Physica C* **185–189**, 86–92 (1991).
8. Mook, H. A. *et al.* *Phys. Rev. Lett.* **70**, 3490–3493 (1993).
9. Fong, H. F. *et al.* *Phys. Rev. Lett.* **75**, 316–319 (1995).
10. Scalapino, D. J. & White, S. R. *Phys. Rev. B* **58**, 8222–8224 (1998).
11. Demler, E. & Zhang, S. C. *Nature* **396**, 733–735 (1998).

Genetics

Engineering a broken heart

Peter J. Scambler

Many human birth defects arise because of chromosomal abnormalities such as deletions or insertions. For example, the most common deletion syndrome in humans — DiGeorge syndrome — results from the deletion of a large region within chromosome 22. Although the DNA sequence of the deleted region has been known for some time, progress in determining which genes in this sequence are critical for development has been slow. This is due to the complexity of human genetics and the limitations of working with human material.

On page 379 of this issue, Lindsay and colleagues¹ report that, using the comparative human-mouse genome maps and chromosome-engineering technology, the DiGeorge syndrome can be modelled — at least in part — in mice. This model will allow researchers to investigate the embryological basis of the malformations seen, and to identify the critical genes. At the same time, their study demonstrates the potential of chromosome engineering for the investigation of birth defects involving chromosome duplications or deletions.

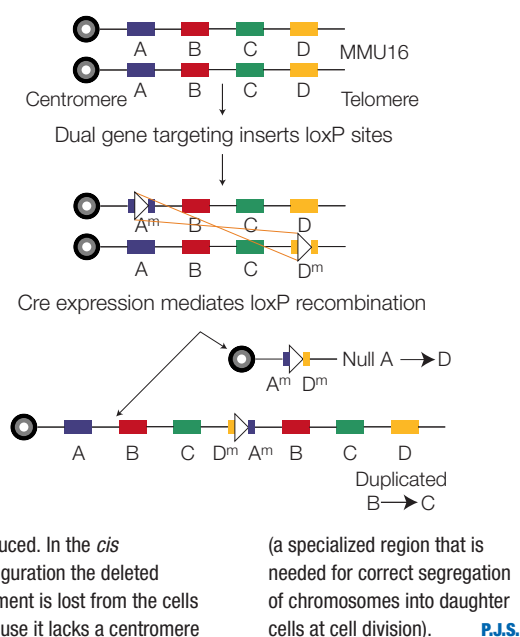
The symptoms of DiGeorge syndrome

typically include congenital heart defects, hypoparathyroidism (low levels of parathyroid hormone), immunodeficiency (secondary to defective development of the thymus), an abnormal facial appearance and learning difficulties. Over 80% of patients have a deleted region on chromosome 22, and this deletion is inherited in about 10% of cases². Some patients also develop psychiatric illness, prompting suggestions³ that a gene (or genes) encoding a predisposition to schizophrenia is also encoded in this region. The deletion commonly affects 3 Mb of DNA, encompassing roughly 30 known genes, and it is thought to occur at a frequency of one in every 4,000 live births.

For normal human development, certain gene products must be present in the correct 'dose'. On the autosomal chromosomes (that is, everything except the X and Y sex chromosomes), one copy of each gene is inherited from each parent, giving a normal dose of two copies — a state known as dizygosity. The presence of extra chromosomal material (say, an increase to three copies of a certain chromosome) or the deletion of chromosomal material (such as reduction to one gene

Box 1 Targeting deletions and duplications

Chromosomal deletions can be engineered by targeting two genes with a bacteriophage loxP sequence. The two loxP sites are inserted at the ends of the desired deletion (at genes A and D in the figure; the superscript 'm' denotes the resulting mutated versions of these genes). The loxP sites may be targeted to the same chromosome (*cis* configuration) or, as in the system used by Lindsay *et al.*¹, on different chromosomes of a homologous pair (*trans* configuration). Recombination between the two loxP sites is catalysed by the Cre recombinase, which is transiently expressed within the embryonic stem cells. In the *trans* configuration a deletion and a reciprocal duplication are



copy; a state known as hemizygosity) can cause congenital abnormalities. The term 'haploinsufficiency' is used when hemizygosity for a gene (or group of genes) causes an abnormality.

For most genes, hemizygosity does not seem to result in haploinsufficiency — that is, we can get by with just one copy of these genes rather than two. But this is not the case in DiGeorge syndrome, and the past decade has seen a search for the one or more 'critical' genes that contribute to this haploinsufficiency. This initially involved comparing the deletions in various patients to work out the shortest region of overlap within which any critical gene should be located. Several years ago the consensus was that, as with other deletion syndromes⁴, haploinsufficiency of just one gene within the 3-Mb deleted region would be responsible for most of the symptoms of DiGeorge syndrome.

But two factors have confused this issue. First, a balanced chromosomal translocation (the switching of DNA from one chromosome to another) was described in one patient with the disease. As no genetic material is lost in such cases, it was assumed that the translocation break-point — where the two chromosomes involved join — directly disrupts a gene implicated in DiGeorge syndrome. Not so. The break-point does not seem to affect any gene directly, and it lies outside the region on chromosome 22 that has been most consistently implicated in the disease⁵. To make matters worse, some deletions found in patients with features of DiGeorge syndrome do not overlap⁶. Two explanations have been offered: either hemizygosity of more than one gene pro-

duces a similar syndrome; or long-range effects on the transcription of one or more critical gene(s) is secondary to some of the chromosome abnormalities.

Faced with the failure of human genetics to uncover which genes are involved in DiGeorge syndrome, Lindsay and colleagues¹ turned to the mouse. The genes deleted on human chromosome 22 map to the same region of mouse chromosome 16, although there has been some shuffling in the order of the genes and compression of the sequence⁷. So, the authors used gene-targeting technology to create both a 1.2-Mb deletion and the corresponding duplication (Box 1) of a section of this presumptive DiGeorge syndrome region.

Lindsay and colleagues found that, although mice with the duplication seemed normal, 10% of mice with one deleted and one normal chromosome (*Dfl*^{-/+} mice) died at birth. Not surprisingly, embryos with a deletion on both copies of chromosome 16 died early during gestation. When the hearts of the *Dfl*^{-/+} embryos were examined just before birth, a quarter had defects typical of those seen in DiGeorge syndrome. In particular, defects in derivatives of the fourth branchial-arch artery were similar to those seen in other experimental models of DiGeorge syndrome⁸. Growth and remodeling of the branchial-arch arteries, which are critical for development of the main arteries and their alignment with the chambers of the heart.

When Lindsay *et al.* bred the mice carrying the duplicated region with mice hemizygous for the deletion, they found that the

duplication removed symptoms of the deletion. This result shows that long-range effects need not be invoked to explain the observed defects. It also indicates that such 'complementation' will be useful in determining which gene (or combination of genes) is dosage sensitive. However, none of the embryos showed any of the other malformations seen in DiGeorge syndrome, and even some of the cardiac defects were restricted. For example, persistence of the truncus arteriosus (which occurs when the common arterial vessel fails to divide into the aorta and pulmonary artery, and is quite common in DiGeorge syndrome) was not observed. Perhaps this reflects a fundamental difference between human and mouse development. Or, more likely, maybe other, background genetic effects modify the observed symptoms, or maybe additional genes need to be deleted to recapitulate the full spectrum of disease. Both of these second possibilities can now be addressed.

Which genes might be critical for DiGeorge syndrome? Human genes from the deleted region on chromosome 22 have been screened for mutations in those cases of DiGeorge syndrome that do not involve deletions. None have been detected. One of the genes that Lindsay and colleagues targeted is called *Ufd1l*, a downstream target of the cardiac transcription factor *Hand2*. On the basis of a small deletion encompassing part of human *UFD1L*, this has been proposed⁹ to be the main gene responsible for DiGeorge syndrome. But mice hemizygous for *Ufd1l* are normal, and an extensive screen for human mutations in *UFD1L* was negative¹⁰. So, genes other than *UFD1L* must be important for the development of DiGeorge syndrome. The location of these genes can now be narrowed by comparing deletions in the mouse. For instance, mice with a 140-kb deletion encompassing seven genes from a region that overlaps *Dfl* have none of the structural abnormalities of DiGeorge syndrome. They do have a behavioural abnormality, though (an increased prepulse inhibition of the startle reflex)¹¹. It will be interesting to investigate the behaviour of the *Dfl* mutants, especially as one of the deleted genes encodes the enzyme proline dehydrogenase, mutations of which can cause behavioural anomalies in the mouse¹².

Over the next year or two, analysis of smaller deletions as well as transgenic complementation experiments will reveal which genes are haploinsufficient in the *Dfl*^{-/+} mice — haploinsufficiency of a single gene could still be responsible for the main features of DiGeorge syndrome. The *Dfl* mutant itself can be used to investigate the developmental processes affected in the disease. And more generally, Lindsay and colleagues' results should be of great encouragement to other groups wishing to model human chromosome abnormalities in the mouse. ■

Peter J. Scambler is at the Molecular Medicine Unit, Institute of Child Health, London WC1N 1EH, UK. e-mail: p.scambler@ich.ucl.ac.uk

1. Lindsay, E. A. *et al.* *Nature* **401**, 379–383 (1999).
2. Emanuel, B. S., Budarf, B. S. & Scambler, P. J. in *Heart Development* (eds Rosenthal, N. & Harvey, R.) 463–478 (Academic, San Diego, 1998).
3. Karayiorgou, M. *et al.* *Proc. Natl Acad. Sci. USA* **92**, 7612–7616 (1995).
4. Li, L. *et al.* *Nature Genet.* **16**, 243–251 (1997).
5. Sutherland, H. F. *et al.* *Am. J. Hum. Genet.* **59**, 23–31 (1996).
6. Rauch, A. *et al.* *Am. J. Hum. Genet.* **64**, 658–666 (1999).
7. Lund, J. *et al.* *Mamm. Genome* **10**, 438–443 (1999).
8. Kirby, M. L. & Waldo, M. L. *Circ. Res.* **77**, 211–215 (1995).
9. Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T. & Srivastava, D. *Science* **283**, 1158–1161 (1999).
10. Wadey, R. *et al.* *Am. J. Hum. Genet.* **65**, 247–249 (1999).
11. Kimber, W. *et al.* *Hum. Mol. Genet.* (in the press).
12. Gogos, J. A. *et al.* *Nature Genet.* **21**, 434–439 (1999).

Climatology

Extremes in the Indian Ocean

David Anderson

The behaviour of the Pacific and Atlantic Oceans is by no means perfectly predictable, but we have some indications of quasi-periodic oscillations in parameters such as water temperature and rainfall. By contrast the Indian Ocean has been an almost complete enigma in this respect. Papers by Webster *et al.*¹ and Saji *et al.*² (pages 356 and 360 of this issue) now furnish evidence of some large-scale interactions between the ocean and atmosphere in the region, and indications of the climatic consequences. But how those regularities may influence or be influenced by other climatic phenomena is far from clear.

This is a topic with a long history. For instance, in 1877 severe drought racked India, Asia and possibly Africa, and resulted in over 20 million estimated deaths from famine and disease³. Such disasters led to attempts to predict the Indian monsoon rainfall, and indeed to the aim of seasonal forecasting in general. One of the greatest proponents of these endeavours was Sir Norman Lockyer, who in addition to his meteorological contributions to science was also co-founder of *Nature* and its first editor. Lockyer's strategy of linking monsoon rainfall to solar cycles proved to be ill-founded. But he did identify the out-of-phase relationship between pressure in the Indian and eastern Pacific Oceans, and so helped in recognizing the El Niño/Southern Oscillation — the interannual (year-to-year) variation between warm and cold phases in the tropical Pacific.

History should not judge Lockyer harshly for failing to provide reliable rainfall predictions, because the Indian Ocean and adjacent region has clung jealously to its secrets. So, for example, although the El Niño/Southern Oscillation has some influence on rainfall around the Indian Ocean, it is not sufficient to explain the observed interannual variability in precipitation. It is against this backdrop that we should view the papers by Webster *et al.*¹ and Saji *et al.*². They describe unusual events in the tropical Indian Ocean in 1997–98 which clearly point

to the existence of a major 'mode' of interaction between the ocean and atmosphere lasting many months.

The tropical Pacific and Atlantic Oceans have certain features in common. The equatorial winds blow from east to west; and the surface temperature is warm in the west but cold in the east, a characteristic associated with a thermocline (a region, about 100 m deep in the equatorial region, where temperature changes rapidly with depth) which slopes up to the east. The Indian Ocean is different. It is warm in the east and slightly cool in the west, features that are associated with winds blowing from the west towards Indonesia. Interannual variability in the Indian Ocean is different too: little coherent behaviour has been identified in it until now, in contrast to the other two oceans where several modes of variability have been evident.

One possible explanation for the apparent lack of variability in the Indian Ocean region is the different relative location of the major precipitation and — through release of latent heat — the associated atmospheric heating regimes. For the Pacific and Atlantic Oceans, heavy precipitation is located to the west, over Indonesia and Amazonia respectively. For the Indian Ocean the heavy pre-

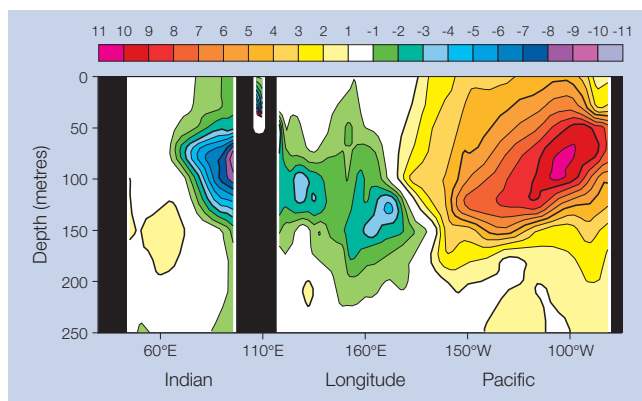
cipitation is in the east, which influences the resulting form of climate variability⁴.

In the second half of 1997, however, there was unusually heavy rainfall over Kenya, Somalia and Ethiopia, and a dramatic — almost two-metre — rise in the level of Lake Victoria⁵. In the east, Indonesia was very dry; there were large-scale forest fires and a thick blanket of smog enveloped the region. Webster *et al.* show that, at this time, anomalies in sea-surface temperature developed in the Indian Ocean (cold in the east, and warm in the west), a state of affairs known as the Indian Ocean dipole. They relate the ocean temperature distribution to increased rainfall over east Africa and reduced rainfall over Indonesia, and to the associated shift in atmospheric heating patterns. Just as for the other equatorial oceans, there is a strong relationship between the east–west temperature gradient along the Equator and the winds on (and to the south of) the Equator, giving the necessary conditions for positive feedback⁶.

One of the characteristics of the tropical oceans is the relatively rapid and efficient channel of communication, between one side of an ocean basin and the other, through Kelvin waves (which propagate eastwards along the Equator at speeds of up to 3 m s⁻¹) and Rossby waves (which are off-equatorial and propagate more slowly westwards). Webster *et al.*¹ argue that these ocean waves, which take a few months to travel round the tropical Indian Ocean, are associated with the intensification, duration and demise of the temperature anomaly in the spring of 1998 (see their Fig. 4 on page 359 for a nice conceptual model of the Indian Ocean mode). Similar waves are implicated in the interannual variability in the Pacific and Atlantic, although their role is not fully understood.

During strong El Niños, such as the one of 1997–98, the Pacific changes into a state similar to that of the 'normal' Indian Ocean — that is, warm in the east, slightly cool in the west. During 1997 the Indian Ocean

Figure 1 Vertical section of the anomalous temperatures along the Equator in the Indian Ocean (left) and Pacific Ocean (right) at the end of October 1997. Only temperatures in the top 250 m are important for climate changes on seasonal timescales. The eastern Indian Ocean was over 10 °C colder than usual at 80 m depth whereas the east Pacific was more than 10 °C warmer than normal. Such large anomalies are very unusual. The changes in the subsurface water give some predictability to El Niño, and may do likewise for the dipole mode in the Indian Ocean identified by Webster *et al.*¹ and Saji *et al.*². (From ref. 10.)



looked like the 'normal' Pacific. Figure 1 shows the behaviour of the subsurface water in the Indian and Pacific Oceans at a time when both oceans were in a highly anomalous state. It is tempting to think that it is no coincidence that a very strong dipole in the Indian Ocean occurred at the same time as one of the largest El Niños on record. But both Webster *et al.*¹ and Saji *et al.*² argue otherwise. They show that the conditions of 1997–98 are not unprecedented and that large amplitudes of the dipole occurred in 1961 (ref. 7), 1967, 1972 and 1994. Dipole variability is present to some degree most of the time but what kicks it into top gear is not clear. There is little statistical relationship between the Indian Ocean dipole and El Niño: the dipole occurs in both El Niño and non-El Niño years, and Saji *et al.* show that it has also been evident even in a weak La Niña year such as 1967 (La Niña, the converse of El Niño, is when the east Pacific is anomalously cold).

Based on six major dipole events in the past 40 years, Saji *et al.*² have constructed a composite picture of the dipole mode showing its spatial structure at different phases. Their Fig. 2 (page 361) highlights the importance of wind and temperature changes first in the Indonesian region, initially to the south of the Equator but then, later, in September–October, westward along the Equator (sometimes, as in 1997, as far as Africa). They show that the dipole mode has a biennial character which may imply a certain predictability from one year to the next. The reliability and so usefulness of such predictions is likely to be limited, however, as we simply don't know how the dipole fits into the larger picture of climate variability and factors such as Himalayan snow cover and chaotic processes that are not directly linked to slow changes in sea-surface temperature.

How the dipole is related to monsoon rainfall is also unclear⁸; of the two biggest dipole events, that of 1961 was associated with the heaviest monsoon in 150 years², whereas in that of 1997 rainfall over India was normal⁹. In fact, Saji *et al.* find no statistical relationship between monsoon rainfall and the dipole. Nonetheless, the observations and analysis reported by Webster *et al.* and Saji *et al.* mean that the Indian Ocean jumps several rungs up the climate ladder of importance. We are still some way from understanding rainfall around the Indian rim, one of Lockyer's original goals, but identification of variability such as the dipole mode is a considerable step forward.

David Anderson is at the European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK.

e-mail: sta@ecmwf.int

1. Webster, P. J., Moore, A. M., Loschnigg, J. P. & Leben, R. R. *Nature* **401**, 356–360 (1999).
2. Saji, N. H., Goswami, B. N., Vinayachandran, P. N. & Yamagata, T. *Nature* **401**, 360–363 (1999).
3. Kininmonth, W. *The 1997–98 El Niño Event: A Scientific and*

Technical Retrospective (World Meteorological Organisation, Geneva, in the press).

4. Anderson, D. & McCreary, J. *J. Atmos. Sci.* **42**, 615–625 (1986).
5. Birkett, C., Murtugudde, R. & Allan, T. *Geophys. Res. Lett.* **26**, 1031–1034 (1999).
6. Bjerknes, J. *Mon. Weath. Rev.* **97**, 163–172 (1969).

7. Reverdin, G., Cadet, D. & Gutzler, D. *Q. J. R. Meteorol. Soc.* **112**, 43–68 (1986).
8. Nicholls, N. *J. Clim.* **8**, 1463–1467 (1995).
9. Bell, G. & Halpert, M. *Bull. Am. Meteorol. Soc.* **79**, S1–S50 (1998).
10. www.ecmwf.int

Signal transduction

Mating, channels and kidney cysts

Scott W. Emmons and Stefan Somlo

Advances in human genetics have meant that the genes mutated in human diseases can be identified exclusively by their location in the genome. But how do we work out the cellular functions of the associated protein products? Reports on pages 383 and 386 of this issue^{1,2} begin to address this problem for two proteins — polycystin-1 (PKD1) and polycystin-2 (PKD2) — that are defective in human autosomal dominant polycystic kidney disease. From their studies of the nematode worm *Caenorhabditis elegans*, Barr and Sternberg² present evidence that homologues of the polycystins act together in a signal-transduction pathway in sensory neurons. Chen *et al.*¹, by contrast, have used an oocyte-expression system in the frog *Xenopus laevis* to show that a homologue of PKD2 is associated with the activity of a cation channel. These results support the

hypothesis that polycystin-related proteins belong to a hitherto unknown class of signal-transduction molecules.

Autosomal dominant polycystic kidney disease affects more than one in 1,000 live births, and is the most common single-gene disorder leading to kidney failure. The children of affected parents have a one in two chance of inheriting the disease, and half of all sufferers need dialysis or a kidney transplant by the age of 60. Mutations in either PKD1 or PKD2 cause almost indistinguishable clinical symptoms. Although the kidney is the most severely affected organ, the disease is systemic and affects the liver, pancreas and cardiovascular and cerebro-vascular systems as well.

The PKD1 protein is a roughly 4,300-amino-acid integral-membrane glycoprotein with a large, amino-terminal extracellular domain and a small, carboxy-terminal

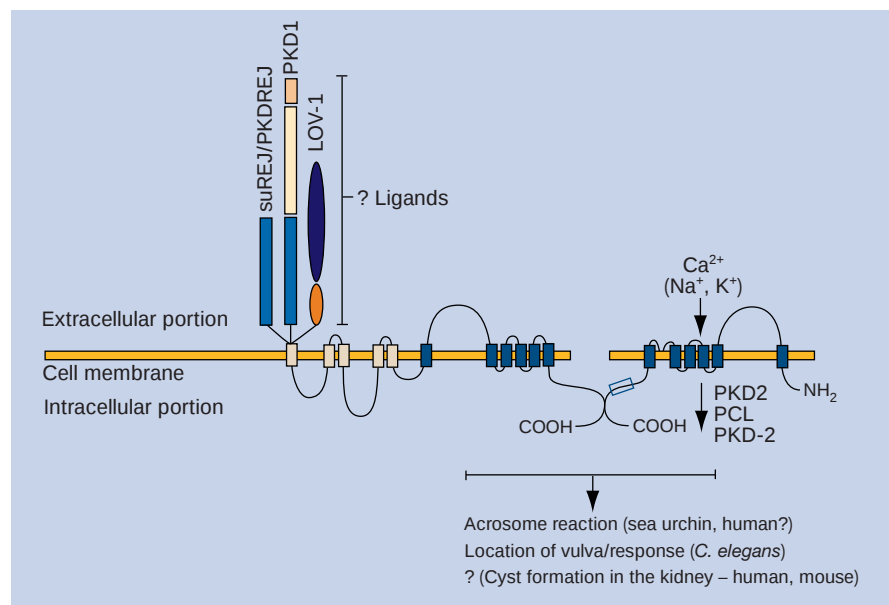


Figure 1 Two components of the proposed polycystin signalling pathway. Proteins related to human PKD1 (refs 3, 4) include the *C. elegans* LOV-1, studied by Barr and Sternberg², the sea-urchin egg jelly receptor (suREJ)⁸ and its human homologue PKDREJ (ref. 9). The large, extracellular domains contain different binding motifs: REJ module (blue); immunoglobulin-like PKD repeats (yellow); leucine-rich repeats and C-lectin-like domain (pale orange); serine/threonine-rich mucin-like domain (purple); ATP/GTP-binding domain (dark orange). The small cytoplasmic tails are thought to mediate interaction with proteins related to human PKD2 (ref. 5), including the polycystin-like (PCL) protein studied by Chen *et al.*¹ and the *C. elegans* PKD-2. PKD2-related proteins are predicted to have six membrane spans (dark blocks) and a cytoplasmic EF-hand (open block). The dark blocks in the PKD1-related protein represent the region of sequence similarity with PKD2-related channels.

cytoplasmic tail^{3,4}. The predicted structure of its domains suggested that it is involved in cell-cell interactions or in interactions with the extracellular matrix. The PKD2 protein has similarities to PKD1, but its topology and domain structure suggest that it might act as a subunit of a cation (perhaps calcium) channel⁵. An important clue to the relationship between PKD1 and PKD2 came from the discovery that they interact directly, suggesting that they act in a common pathway^{6,7}. The possibility of a widespread function of related pathways was first suggested by the discovery of a structural relationship between PKD1 and a membrane receptor in sea-urchin sperm. This receptor mediates the acrosome reaction — an ion-channel-regulated membrane fusion event that is necessary for fertilization⁸. Human testes express a similar polycystin-related protein⁹.

Now, quite unexpectedly, Barr and Sternberg² find that a mutation in *C. elegans*, which gives rise to males that are defective in mating behaviour, lies in a gene called *lov-1* (for 'location of vulva') — the worm homologue of human *PKD1*. Since research on *C. elegans* was initiated by Sydney Brenner in the late 1960s, work on this animal has, in part, been directed at understanding the genetic basis of development and the function of its nervous system. Male mating — in which males seek out the hermaphrodite partner and copulate with her — is probably the most complex behaviour shown by *C. elegans*.

Sternberg's laboratory had previously defined the six sub-steps of the stereotyped copulatory sequence, correlated these sub-steps with the function of individual neurons, and isolated behavioural mutants¹⁰. One of the sub-steps is to locate the vulva. As well as being unable to execute this step efficiently, *lov-1* mutant males are also defective in the first sub-step, termed 'response'. Response and vulva location depend on two types of male sensory structure. The first is a set of nine pairs of rays, which project out of the tail on each side. The second is a hardened cuticular structure called the hook, which contains two sensory neurons.

Knowing that PKD1 and PKD2 interact, Barr and Sternberg next used the recently completed *C. elegans* genome sequence to isolate *pkd-2*, the worm homologue of human *PKD2*. They then studied the expression patterns of both *lov-1* and *pkd-2*, and found that promoter sequences of both genes cause reporter genes to be expressed in the rays and the hook sensory neurons required for response and vulva location. Arguing (from a variety of evidence) that the defect in the *lov-1* mutant is not developmental, the authors concluded that the *LOV-1* and *PKD-2* proteins are involved in chemosensory or mechanosensory signal transduction in sensory neurons.

By contrast, Chen *et al.*¹ used cell-expres-

sion and electrophysiological approaches to examine the potential channel function of a polycystin-related protein. This protein, called PCL (polycystin-like), had been identified in the human expressed-sequence-tag database by its sequence similarity with that of PKD2. Although its function was not known, the authors knew that PCL, like PKD2, has the structural fingerprint of a cation-channel subunit related to a number of families (the transient receptor potential calcium channel and voltage-gated calcium-, sodium- and potassium-channel families). The PCL and PKD2 proteins both contain a calcium-binding EF-hand domain that may help to regulate channel activity.

Chen *et al.* expressed PCL in *Xenopus* oocytes by microinjecting synthetic messenger RNA for the protein. They then studied its channel properties using the two-micro-electrode voltage clamp and patch-clamp techniques. The authors found that PCL is a non-selective cation channel that is permeable to sodium, potassium and calcium. Its calcium permeability is about five-fold higher than that of sodium, and calcium modulates the channel's activity. However, Chen *et al.* could not determine whether binding at the EF-hand domain is responsible for this calcium regulation. The high structural similarity between PCL and PKD2 provides indirect evidence that PKD2 is also a cation-channel subunit.

These data support the hypothesis that PKD1-related proteins act as receptors that regulate the activity of channels containing PKD2-related proteins (Fig. 1). The two proteins are part of a conserved signalling mechanism in which the translocation of ions acts as a second messenger. But the diversity of the processes in which this signalling mechanism seems to be involved highlights the remaining questions. What are the specific molecular cues that activate these pathways? What are their downstream effectors? And what additional factors are responsible for adapting this mechanism to the unique requirements of each tissue? ■

Scott W. Emmons is in the Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. e-mail: emmons@acom.yu.edu

Stefan Somlo is in the Section of Nephrology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520-8029, USA. e-mail: stefan.somlo@yale.edu

1. Chen, X. Z. *et al.* *Nature* **401**, 383–386 (1999).
2. Barr, M. M. & Sternberg, P. W. *Nature* **401**, 386–389 (1999).
3. Hughes, J. *et al.* *Nature Genet.* **10**, 151–160 (1995).
4. The International Polycystic Kidney Disease Consortium *Cell* **81**, 289–298 (1995).
5. Mochizuki, T. *et al.* *Science* **272**, 1339–1342 (1996).
6. Qian, F. *et al.* *Nature Genet.* **16**, 179–183 (1997).
7. Tsiokas, L., Kim, E., Arnould, T., Sukhatme, V. P. & Walz, G. *Proc. Natl Acad. Sci. USA* **94**, 6965–6970 (1997).
8. Moy, G. W. *et al.* *J. Cell Biol.* **133**, 809–817 (1996).
9. Hughes, J., Ward, C. J., Aspinwall, R., Butler, R. & Harris, P. C. *Hum. Mol. Genet.* **8**, 543–549 (1999).
10. Liu, K. S. & Sternberg, P. W. *Neuron* **14**, 79–89 (1995).

Daedalus

Fertile competition

The claim has been made that sperm counts in the West are declining alarmingly. A typical ejaculate might contain 100 million sperm; since only one is required to do the job, a reduction to (say) 50 million may not seem obviously critical. But human fertilization is chancy at best. Even trying hard, a couple can easily take many months to conceive. One explanation is that most sperm are infertile. Their job is to ward off or discourage rival sperm. In effect, they act as a large screen of warships escorting a small, crucial convoy of freighters.

Daedalus argues that both freighters and warships will put on a mighty spurt if challenged by a rival fleet. There is some evidence that a man with a sexual rival generates more sperm than he would do otherwise; but Daedalus reckons that the speed, efficiency and pugnacity of his sperm must rise as well. In many species sperm compete chemically, by putting out toxins or antigens against their rivals. Indeed, Daedalus once proposed to use human seminal toxins in a natural spermicidal contraceptive. He now has a converse strategy. DREADCO biochemists are studying human seminal toxins in the hope of developing a spermal 'vaccine'. It will be a derivative of such a toxin, modified just enough to be harmless, but still sensed as a deadly threat by sperm encountering it. Spurred by this challenge, they will drive towards the ovum with extra speed and energy. This ingenious 'conceptive' will be welcomed by couples trying hard to have children. It will boost their chances greatly.

But Daedalus goes further. The sperm in a given ejaculate must be immune to their own toxin. They should even tolerate quite well the toxin of a close genetic relative carrying many of the same genes. But toxin from a genetic stranger must be a terrible threat. The DREADCO team are therefore mixing semen samples from different types and races of men, and studying their competition under the microscope. They will then plot the semen donors on a map such that the more fiercely antagonistic the sperm of any two donors, the further apart they are on the map.

The resulting human distribution will be far more fundamental than one based (say) on blood groups or pigmentation. It will reveal the classes of mankind as sensed by genetics itself. It should powerfully illuminate the stages by which we emerged from Africa, and our diversification since then.

David Jones

Discovery of tetraploidy in a mammal

The red viscacha rat is unaffected by having double the usual number of chromosomes.

Polyploidy, or having more than a pair of each type of chromosome, is considered to be unlikely in mammals because it would disrupt the mechanism of dosage compensation that normally inactivates one X chromosome in females¹. Also, any imbalance in chromosome number should affect the normal developmental processes and therefore constitute an evolutionary end, as in triploid humans².

Nevertheless, genetic evidence indicates that the red viscacha rat, *Tympanoctomys barrerae* (Octodontidae)³, is tetraploid. *T. barrerae*, which is a highly specialized desert rodent⁴, has the largest chromosome complement in mammals⁵ with a single XY sex-chromosome system (Fig. 1). Although little is known about meiotic chromosome pairing in *T. barrerae*, the constant diploid number found in 13 specimens examined at three localities in Argentina indicates that segregation occurs normally.

In mammals without excessive constitutive heterochromatin in their chromosomal complement, nuclear DNA content is relatively constant, ranging from 6 to 8 picograms of DNA (mean, 6.3 pg)⁶. Analysis of somatic tissues by flow cytometry indicates that *T. barrerae* has a genome of 16.8 pg, which is twice the DNA content of its closest relatives (for example, 8.2 pg and 7.6 pg for *Octodontomys gliroides* and *Octomys mimax*, respectively) and of most other mammals.

The huge genome of *T. barrerae* cannot be explained by an increased amount of heterochromatin because it is restricted to the centromere region of chromosomes. Furthermore, the diploid number in *T. barrerae* is less than that would be expected ($2N=112$) from a simple duplication of the karyotype seen in its closest relatives, indicating that chromosomal material has been eliminated.

The size of the sperm head in mammals depends on its DNA content, as exemplified by the abnormally large and diploid sperm of rabbits and bulls⁷. The enormous broad and spatulate sperm head of *T. barrerae* (Fig. 2) is $14.1 (\pm 0.6)$ by $13.4 (\pm 0.5)$ micrometres in size, and gametic genome estimates (*Octomys mimax*, 4.32 ± 0.25 pg; *Spalacopus cyanus*, 3.54 ± 0.27 pg; *T. barrerae*, 9.2 ± 0.72 pg).

As would be expected in a tetraploid organism, significantly larger somatic-cell diameters were recorded. A highly significant paired *t*-test ($P < 0.001$) supported the idea that there was a difference in the maximum diameter of liver cells ($N=150$)

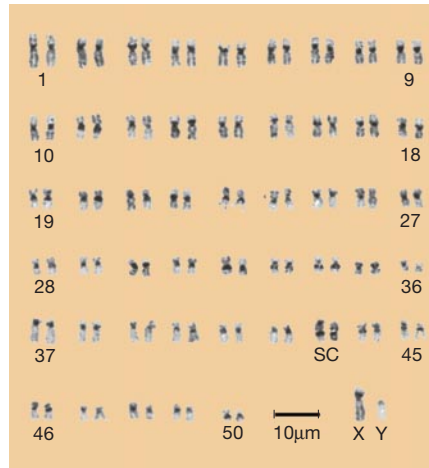


Figure 1 C-banded karyotype of a male *Tympanoctomys barrerae* from Mendoza, Argentina ($4N=102$; fundamental number of autosomal arms is 200). The karyotype includes 36 pairs of metacentric to submetacentric chromosomes and 14 pairs of subtelocentric autosomes. The X chromosome is the largest element, present in two copies in females. The Y chromosome, which is medium-sized, is the only acrocentric element of the karyotype. Note that the marker chromosome pair of the family Octodontidae (secondary constriction, SC), which has a band of interstitial heterochromatin, is present in only two copies, not four, as would be expected following a perfect duplication.

between *T. barrerae* ($26.1 \pm 3.8 \mu\text{m}$) and *Abrocoma bennetti* ($22.3 \pm 3.5 \mu\text{m}$) or *Octodon lunatus* ($21.3 \pm 2.7 \mu\text{m}$).

The müllerian explanation for the rarity of tetraploidy among animals with XY sex-chromosome systems emphasizes inviability or sterility resulting from an altered dosage-sensitive regulatory mechanism⁸. Sex determination in mammals depends on the testis-determining locus, but in the germ line it is affected by dosage of X-chromosome-linked genes⁹. The fact that polyploid cells have more than one active X chromosome indicates that the signal to initiate inactivation through heterochromatinization may be dependent on the critical X-to-autosomal chromosome ratio.

If there is an evolutionary elimination of chromosomes, as we propose, then the X-to-autosome ratio should not depend on the whole set of autosomes, but rather on some critical autosomes needed in only one copy of the gamete. Thus, the disomic condition of the X chromosome in female *T. barrerae* is either sufficient or possibly the only means by which the tetraploid can escape arrested gonadal development. This idea is supported by the inviability or infertility that affects human triploids, who do not have enough diploid cells to undergo meiosis¹⁰.

Investigation of the chromosomal architecture (by studying meiotic pairing and *in*

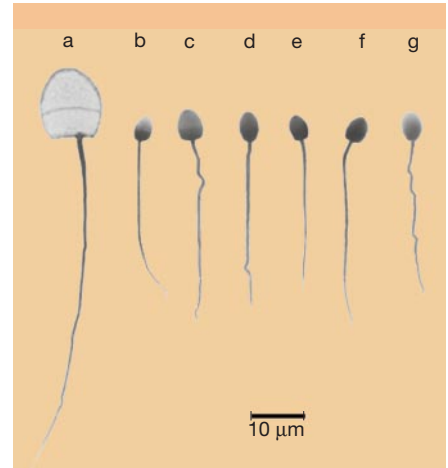


Figure 2 Bright-field photomicrograph of sperm cells from different genera in the families Octodontidae and Abrocomidae. **a**, *Tympanoctomys barrerae*; **b**, *Aconaemys fuscus*; **c**, *Octodon degus*; **d**, *Spalacopus cyanus*; **e**, *Octomys mimax*; **f**, *Octodontomys gliroides*; **g**, *Abrocoma bennetti*.

situ hybridization) and genome organization (by looking for single-dose genes and clusters of linked duplicated genes) of *T. barrerae* should enable this tetraploidy in *T. barrerae* to be verified and will provide new insights into the mechanisms underlying dosage compensation, sex determination, regulation of gene expression and development in mammals. The role of whole-genome duplication in triggering evolutionary novelty is exemplified by this unique tetraploid mammal³.

Milton H. Gallardo*, **J. W. Bickham†**,
R. L. Honeycutt‡, **R. A. Ojeda‡**,
N. Köhler*

*Instituto de Ecología y Evolución,
Universidad Austral de Chile,
Casilla 567, Valdivia, Chile
e-mail: mgallard@uach.cl

†Department of Wildlife and Fisheries Sciences,
Texas A&M University, College Station,
Texas 77843, USA

‡Instituto Argentino de Investigación de Zonas
Áridas, Cricyt, GIB Conicet Casilla de Correo 507,
5500 Mendoza, Argentina

- Orr, H. A. *Am. Nat.* **136**, 759–770 (1990).
- Niebuhr, E. *Humangenetik* **21**, 103–125 (1974).
- Gallardo, M. H. in *Chromosomes Today* Vol 12 (eds Henriques-Gil, N., Parker, J. S & Puertas, M. J.) 347–365 (Chapman & Hall, London, 1997).
- Ojeda, R. A. *et al. J. Arid Environ.* **41**, 443–452 (1999).
- Contreras, L. C., Torres-Mura, J. C. & Spotorno, A. E. *Experientia* **46**, 506–508 (1990).
- Vinogradov, A. E. *Cytometry* **31**, 100–109 (1998).
- Ferrari, M. R., Spirito, S. E., Giuliano, S. M. & Fernández, H. A. *Andrologia* **30**, 85–89 (1998).
- Müller, H. *Am. Nat.* **59**, 346–353 (1925).
- Parkhurst, S. M. & Meneely, P. M. *Science* **264**, 924–932 (1994).
- Ohno, S., Kittrell, W. A., Christian, L. C., Stenius, C. & Witts, G. A. *Cytogenetics* **2**, 42–49 (1963).

Putting plaids in perspective

Farell has described¹ a “new view” of the stereo-matching problem. He claimed that occlusion relationships in natural scenes introduce interocular positional shifts in all two-dimensional (2D) directions, which putatively demonstrates that “horizontal disparity is not a reliable cue to depth” in natural scenes. Farell argued that the visual system must therefore search locally in all 2D directions to establish correspondence.

There is both a geometric and a perceptual component to Farell's claim. The geometric component asserts that occlusion causes corresponding surface regions to be shifted in all 2D directions. However, as the portions of the contour behind Farell's apertures that cannot be superimposed by a horizontal shift are seen by only one eye^{2,3}, they are not disparities (which by definition require matches). Thus, from a geometric perspective, ‘aperture disparities’ do not exist. The geometric consequence of stereoscopic occlusion is to generate monocular features in addition to horizontal disparity²⁻⁷; it does not generate local disparities in all 2D directions, as Farell claims.

The perceptual component depends on whether the visual system actually computes aperture disparities (even though these would be geometrically incorrect). We have shown that the visual system uses the non-horizontal shifts of contour junctions to decompose contours into corresponding and non-corresponding (monocular) contour segments, which implies that, for most (if not all) occlusion configurations, it does not. This is true even in the pattern used by Farell to demonstrate his ‘aperture disparities’ (Fig. 1).

But as Farell used a very different stimulus, his suggestion that aperture disparities are calculated may stem from this difference. He added two sinusoidal gratings with different orientations to create stereo-

scopic plaid patterns, and introduced disparity by horizontally shifting one grating in one of the binocular images. Depending on the orientations of the two gratings, this manipulation would shift the plaid intersections not just horizontally, but in any 2D direction. Farell found that these stereo plaids appeared in a single-depth plane, consistent with the horizontal disparity of the plaid intersections, rather than splitting into two gratings in depth.

It is unclear how this result relates to aperture disparities, as coherence does not occur for the intersections formed by stereoscopic occlusion junctions^{2,3}. The fact that perceived depth was predicted by the horizontal disparity of the plaid's intersections indicates that horizontal disparity is sufficient to understand this result.

Farell reasoned that the perceived depth of the plaids could arise from a second stage of stereo processing that integrates the disparities of the 1D components (gratings) into a coherent surface, so he did an adaptation experiment to determine whether the disparities of the plaid's 1D components could influence a post-adaptation test stimulus. The results provided evidence for facilitation by 1D adapters, but also revealed inhibition by the 2D adapters at the disparity of the test stimulus. Thus, as in previous investigations into this problem^{8,9}, Farell's results are ambiguous about the nature of the matching primitives used to establish binocular correspondence.

Farell's claims about stereoscopic occlusion are grounded on an incorrect understanding of stereo-occlusion geometry, so it is not surprising that he was unable to find unambiguous evidence that the visual system contains mechanisms to compute what he has (mis)labelled “aperture disparities”.

Barton L. Anderson

*Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 01239, USA*

1. Farell, B. *Nature* **395**, 689–693 (1998).
2. Anderson, B. L. *Nature* **367**, 365–367 (1994).
3. Anderson, B. L. *Perception* **26**, 419–453 (1997).

4. Lawson, R. B., Cowan, E., Gibbs, T. D. & Whitmore, C. D. *J. Exp. Psychol.* **103**, 1142–1146 (1974).
5. Gillam, B. & Borsting, E. *Perception* **17**, 603–608 (1988).
6. Nakayama, K. & Shimojo, S. *Vision Res.* **30**, 811–825 (1990).
7. Gillam, B., Blackburn, S. & Nakayama, K. *Vision Res.* **39**, 493–502 (1999).
8. Morgan, M. J. & Castet, E. *Vision Res.* **37**, 2737–2744 (1997).
9. Hibbard, P. B. & Langley, K. *Vision Res.* **38**, 1073–1084 (1998).

Farell replies — When we look at Anderson's Fig. 1a, or at just about anything else, we see surfaces and boundaries, contours and intersections, edges and angles. These complex, interpreted image features might well be the stimulus elements that stereo mechanisms analyse to recover depth¹⁻⁴. But are they?

To answer this question, I proposed a two-stage stereo framework beginning with the matching of simple orientated components that are not elements of our conscious experience, but are standard elements of psychophysical and physiological theorizing. In the light of Anderson's interpretation, in contrast, complex image features such as surfaces, boundaries and intersections become the inputs to perceptual processing, rather than the outcomes of it.

One interpretative difference concerns definitions. In Anderson's definition, only points on the left and right retinal images that correspond to the same point in 3D space have a disparity. Looking at any stereogram, we see why this objectivist definition, though appropriate when studying optics, can be misleading when studying perception. The left and right halves of a stereogram, being spatially separated, have no point in common, so by this definition stereograms cannot have disparities. This, of course, will not do. But are stereograms not laboratory artifices and therefore weak tests of a classic definition? Even if they were, there are naturalistic Wheatstone viewing conditions in which we see stereoscopic depth from non-veridical matches, such as the ‘double-nail’ and ‘wallpaper’ illusions⁵ — and, I believe, aperture viewing — all of which defy Anderson's definition.

My Fig. 1 (ref. 6) addressed whether non-veridical matches occur when we look through apertures, which function like stereoscopes to give each eye a view of a different part of the world. If observers do make these non-veridical matches, then stereo correspondence must be a 2D, not a 1D, matching process. These matches can be studied using either depth-coherent patterns (for example, sinewave plaids) or segregated patterns (for example, Fig. 1a and most squarewave plaids; see my Fig. 2 (ref. 6)). But coherent plaids allow us to dissociate easily the disparities of 1D components and 2D features and thereby to isolate stereo-matching primitives, which is why I used them in preference to non-cohering patterns.

As Anderson points out, my Fig. 2 (ref. 6) indicates that depth polarity discrimination depends on the horizontal component of

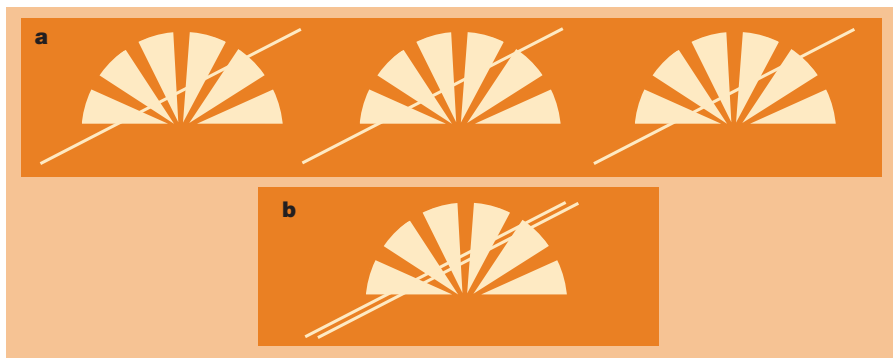


Figure 1 Sunburst stereogram. **a**, A variation of Farell's stereogram that putatively demonstrates that occlusion geometry can cause disparity to be generated in all 2D directions. Cross-fusers should fuse the left two images, divergers the right two images. When fused, the black contour appears behind the slits in a single-depth plane consistent with the horizontal disparity of the line. **b**, The views of the left and right eye are superimposed so the disparity of the apertures is zero. Although the contour segments within the apertures appear shifted parallel to the aperture boundaries, only portions of the contour that can be superimposed by a horizontal shift are corresponding sections of the contour.

2D feature disparities. And he is right that this result is neutral about whether we match 2D features or 1D components, which is why I ran the adaptation experiment. Anderson infers that there is an ambiguity between some marginal inhibitory effects, which suggest contrast adaptation, and robust facilitatory effects, which imply a component analysis of disparity. Contrast adaptation, which is mostly non-selective for depth⁷, could display a selective component at the depth of visible features. But because the facilitatory effect of disparity adaptation occurs at a different depth, the selective component of contrast adaptation, like masking⁸, must occur after stereo matching. This reinforces a component analysis of disparity. Still, the main result remains the improved performance after adaptation at the components' disparity, which other models do not explain.

As Anderson says, depth coherence is not usually seen between occluding and occluded surfaces, and seems absent from real-world scenes. Coherence is easily dismissed as an illusion, as a failure to segregate the visual world into distinct objects, which could be hazardous if it were more than just a laboratory curiosity. If 1D components are the stereo-matching primitives, we will have to flip this view on its head.

A broadband object in depth contains 1D components with phase disparities that potentially signal an assortment of depths⁸. Visual systems do not process these components as independent stimulus elements, but combine them into visible features, such as surfaces, boundaries, edges and intersections, and combine their disparities into a particular (coherent) depth, just as they combine the components of sinusoidal plaids into 2D features and give them coherent depth. Coherence is the rule and is an integral part of object perception, not exceptional or illusory. Coherence is also compatible with, and necessary for, depth segregation: sets of components that cohere around different phase-congruent disparities will automatically segregate in depth.

My conclusions applied to stereoscopic matching generally, not just to matching in apertures or transparent scenes. As Anderson notes, apertures and other partial occluders typically result in failures to match. I did not examine failures to match, which do not account for the study's empirical findings: depth coherence, depth reversal and disparity adaptation of 1D components.

Elsewhere, Anderson and others have shown that unpaired image regions resulting from partial occlusions do indeed influence perceived depth (see his references). Like other binocular information sources, these unpaired regions ('visibility disparities') can be detected only through a matching process. We must include visibility disparities among the many cues available

from stereo matching, including disparities of spatial frequency and orientation, as well as position or phase, all of which bear on a scene's 3D spatial layout and yet are potentially conflicting. Reliable perception of the third dimension requires that evidence from all these sources be combined at a second-stage site. My study is therefore not incompatible with Anderson's previous work: partial occlusions should fit comfortably into the two-stage framework.

Bart Farell

Institute for Sensory Research, Syracuse University,
Syracuse, New York 13244-5290, USA
and Center for Vision Research,
SUNY Health Science Center,
Syracuse, New York 13210-2375, USA
e-mail: bart_farell@isr.syr.edu

1. Ogle, K. N. (1962) in *The Eye*, Vol. 4. *Visual Optics and the Optical Space Sense* (ed. Davson, H.) 271–324 (Academic, New York, 1962).
2. Bishop, P. O. in *The Neurosciences: Second Study Program* (ed. Schmitt, F. O.) 471–485 (Rockefeller Univ. Press, New York, 1970).
3. Julesz, B. *Foundations of Cyclopean Perception* (Chicago Univ. Press, 1971).
4. Howard, I. P. *Human Visual Orientation* (Wiley, New York, 1982).
5. Howard, I. P. & Rogers, B. J. *Binocular Vision and Stereopsis* (Oxford Univ. Press, New York, 1995).
6. Farell, B. *Nature* **395**, 689–693 (1998).
7. Blakemore, C. & Hague, B. J. *Physiol.* **225**, 437–455 (1972).
8. McKee, S. P., Bravo, M. J., Taylor, D. G. & Legge, G. E. *Vision Res.* **34**, 1047–1060 (1994).

Nutrition

Effect of vegetables on bone metabolism

Fractures caused by osteoporosis are a major burden to health care¹. Attempts to prevent osteoporosis through diet have had little success: calcium consumed in dairy products has only a small effect on the risk of hip fractures², and soy, a rich source of phytoestrogens that has been proposed as an alternative to oestrogen treatment³, has not yet been shown to be effective in humans. Here we show that a variety of salads, herbs and cooked vegetables that are common in the human diet can alter bone metabolism in rats.

The consumption by rats of onion, for example, increases their bone mass: in male rats fed 1 g dry onion per rat per day for 4 weeks, bone mineral content⁴ increased by $17.7 \pm 6.4\%$ ($P < 0.05$; $n = 6$), mean cortical thickness increased by $14.8 \pm 7.6\%$, and the mineral density of trabecular bone increased by $13.5 \pm 3.1\%$ ($P < 0.05$) relative to controls.

Figure 1 shows that 14 vegetables eaten by humans can significantly inhibit bone resorption in the rat. A mixture of 500 mg each of onion and Italian parsley (Fig. 1, lane 35), and a mixture of lettuce, tomato, cucumber, arrugula (rocket), onion, garlic, wild garlic, common parsley, Italian parsley and dill (100 mg of each daily) significantly

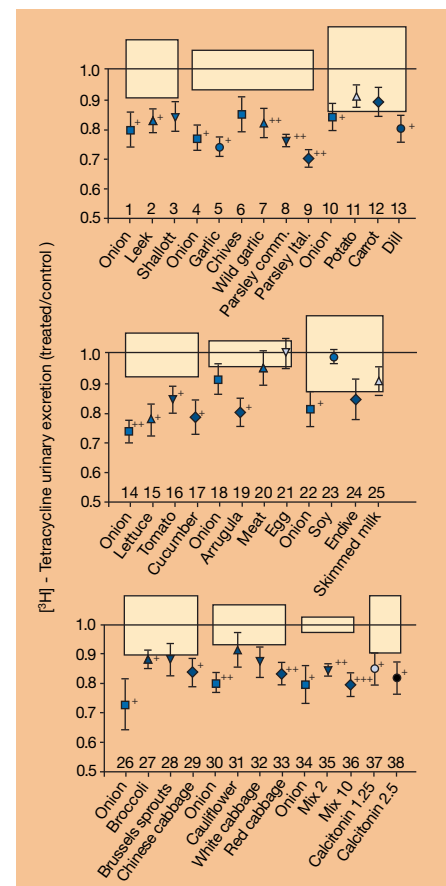


Figure 1 Effect of foodstuffs and the hormone calcitonin on bone resorption as assessed by the urinary excretion of previously administered radiolabelled tetracycline^{9–10}. Data are plotted as the ratio of treated/untreated control ($\pm$ s.e.m.; $n = 5$ per group) over 10 days. The 95% confidence interval of the untreated control groups ($n = 5–6$; 10 experiments) is shaded. Onion was used as a positive control for all foodstuffs. All rats received the same total daily amount of food, including 1 g of the dried test foodstuff. Fresh foods were air- or freeze-dried and ground. Calcitonin was injected daily at doses of 1.25 or 2.5 IU per kg body mass at the optimal time⁹. *Cooked before drying; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Details of methods and materials can be obtained from R.C.M.

inhibited bone resorption (lane 36), indicating that the effect was additive. There was no inhibition by soybeans at the same dose, or by foodstuffs of animal origin (lanes 20, 21, 25), and even milk powder had no significant effect, despite its 1.29% calcium content. The mean 20% inhibition by 1 g onion per day is slightly higher than the effect of calcitonin at doses (per kg body weight) used to treat postmenopausal osteoporosis⁵ (lanes 37, 38).

Because osteoporosis in humans occurs most frequently in postmenopausal women, we studied the ovariectomized rat^{6,7} as a model. Although bone resorption in these rats increased by $32 \pm 3\%$ ($P < 0.001$) compared with sham-operated animals, this was inhibited by onion (30–1,500 mg per day) in a dose-dependent manner: at the highest dose, resorption decreased by $25 \pm 4\%$ ($P < 0.01$). Onion therefore inhibits

resorption not only in male rats, but also in females in which bone resorption is stimulated by oestrogen withdrawal.

Food intake in the rat is associated with a large and rapid increase in bone resorption that occurs in a diurnal rhythm^{8,9}, which is affected differentially by inhibitors of bone resorption, such as calcitonin and the bisphosphonate alendronate, because of their different mechanisms of action⁹. The effect of alendronate is maximal after about three days, whereas calcitonin works within a few hours to inhibit the height of the peak only during the diurnal rhythm; oestrogen treatment works only after two days. Peak height is inhibited similarly by calcitonin and onion, the maximal effect being attained 6 to 12 hours after ingestion. The effect of onion on the acute regulation of bone resorption therefore resembles that of calcitonin.

Our results indicate that several common vegetables in the human diet alter bone metabolism in the rat. If this also happens in humans, then including an appropriate amount of these vegetables in the daily diet could be an effective and inexpensive way to decrease the incidence of osteoporosis.

Roman C. Mühlbauer, Feng Li

Bone Biology Group,

Department of Clinical Research,

University of Bern, Murtenstrasse 35,

3010 Bern, Switzerland

- White, P. & Compston, J. *Osteoporosis: Clinical and Commercial Perspectives* (Carter Mill International, London, 1998).
- Kanis, J. A. *Bone* **24**, 279–290 (1999).
- Anderson, J. J. B. & Garner, S. C. *Nutr. Res.* **17**, 1617–1632 (1997).
- Li, F. & Mühlbauer, R. C. *J. Bone Miner. Res.* **14**, 1457–1465 (1999).
- Silverman, S. L. *et al.* *Bone* **23**, 1108–1108 (1998).
- Wronski, T. J., Walsh, C. C. & Ignaszewski, L. A. *Bone* **7**, 119–123 (1986).
- Cecchini, M. G., Fleisch, H. & Mühlbauer, R. C. *Calcif. Tissue Int.* **61**, S9–S11 (1997).
- Mühlbauer, R. C. & Fleisch, H. *Am. J. Physiol.* **259**, R679–R689 (1990).
- Mühlbauer, R. C. & Fleisch, H. *J. Clin. Invest.* **95**, 1933–1940 (1995).
- Antic, V. N., Fleisch, H. & Mühlbauer, R. C. *Calcif. Tissue Int.* **58**, 443–448 (1996).

Genetic recombination

Intron size and natural selection

Intron sizes vary widely among different genes and among homologous genes of different species. The distribution of intron sizes may be maintained in a steady state, reflecting the processes of insertion and deletion of gene sequences, or it may be that the distribution is constrained by natural selection^{1–3}. If intron size is governed by natural selection, there should be a statistical association between this size and the rate of recombination per map unit of the genome, assuming that natural selection is less effective in genomic regions of low

recombination^{4–6}. Here we show that larger introns of *Drosophila melanogaster* occur preferentially in regions of low recombination, which is consistent with large introns having a deleterious effect. The association is significant ($P=0.001$, linear regression), despite the fact that no effort was made to stratify the data by other factors that affect intron size, such as the size of the associated coding region⁷ or the presence of regulatory sequences inside the intron.

The mechanisms by which splice sites are selected and the evolutionary patterns are different in 'small' (up to 80 base pairs) and 'large' (more than 80 base pairs) introns in *Drosophila*^{3,8–10}. Size-class transitions in homologous introns are frequent (about 20%) between *D. melanogaster* and *D. virilis* or *D. pseudoobscura* but rare among most closely related species^{3,11}. We investigated the relation between intron size and recombination rate by analysing each intron type separately (Fig. 1a), and found that recombination occurs in small introns on average at 2.45 ± 0.04 centimorgans per million base pairs (cM Mb^{-1}), whereas the value for large introns is 2.20 ± 0.05 cM Mb^{-1} , a difference that is significant at the 0.001 level (by analysis of variance).

Within large introns, there is no significant association between recombination rate and intron size ($b = -0.012$, $P > 0.2$), so whatever causes this relation must act equally against all large introns, irrespective of their absolute size. In small introns, there is a significant, positive association between recombination rate and intron size ($b = +0.002$, $P = 0.02$), indicating that very small introns are also deleterious: they tend to occur in regions of low recombination, where natural selection is less efficient.

The average rate of recombination for introns of less than 60 base pairs is 2.26 ± 0.07 cM Mb^{-1} , whereas the rate for introns of 60 to 80 base pairs is 2.56 ± 0.06 cM Mb^{-1} , a difference significant at the 0.001 level (Fig. 1b). These findings are not surprising, as it is known that very small introns do not splice well⁸.

Another explanation for these results is that intron size is a neutral trait, with recombination alone generating the association between recombination rate and intron size by its effect on DNA insertion and deletion processes. However, very small and large introns are both associated with regions of low recombination, excluding not only this possibility, but also any other that does not involve the direct action of natural selection on intron size. Fewer than 2% of the large introns in our sample show any sequence similarity to transposable elements, so our results cannot be explained by the accumulation of these elements in regions of low recombination^{3,12}.

Intron size usually changes through small deletions and insertions of DNA^{2,3,13}.

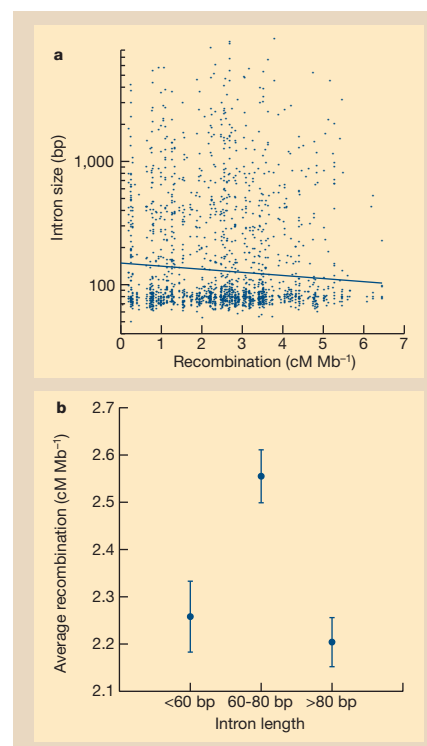


Figure 1 Association between intron size and recombination rate.

a, Relation between local recombination rate and intron size over 1,817 introns (from 590 GenBank accessions) in expressed genes of *Drosophila melanogaster* (linear regression on log-transformed size: $b = -0.026$, $P = 0.001$, $r^2 = 0.006$). Recombination rates were estimated as described⁵. **b**, Average rates of recombination ($\pm$ one standard error) for introns of less than 60, 60–80, and more than 80 base pairs (bp); n is 370, 665 and 782, respectively.

Our results suggest that natural selection in *Drosophila* acts on the variants generated by these processes, selecting against large introns and very short introns, as both tend to occur in regions of low recombination where selection is less effective.

Antonio Bernardo Carvalho*†,

Andrew G. Clark†

**Departamento de Genética,*

Universidade Federal do Rio de Janeiro,

Rio de Janeiro, Brazil

†Institute of Molecular Evolutionary Genetics,

208 Mueller Laboratory,

Pennsylvania State University,

University Park, Pennsylvania 16802, USA

- Hughes, A. L. & Hughes, M. K. *Nature* **377**, 391 (1995).
- Ogata, H., Fujibuchi, W. & Kanehisa, M. *FEBS Lett.* **390**, 99–103 (1996).
- Moriyama, E. N., Petrov, D. A. & Hartl, D. L. *Mol. Biol. Evol.* **15**, 770–773 (1998).
- Hill, W. G. & Robertson, A. *Genet. Res.* **8**, 269–294 (1966).
- Kliman, R. M. & Hey, J. *Mol. Biol. Evol.* **10**, 1239–1258 (1993).
- Cameron, J. M., Kreitman, M. & Aguadé, M. *Genetics* **151**, 239–249 (1999).
- Moroni, G. in *An Atlas of Drosophila Genes* (ed. Moroni, G.) 319–332 (Oxford Univ. Press, 1993).
- Mount, S. M. *et al.* *Nucleic Acid Res.* **20**, 4255–4262 (1992).
- Guo, M. & Mount, S. M. *J. Mol. Biol.* **253**, 426–437 (1995).
- Kennedy, C. F. & Berget, S. M. *Mol. Cell. Biol.* **17**, 2774–2780 (1997).
- Stephan, W. *et al.* *Genetics* **138**, 135–143 (1994).
- Snieszko, P. D. & Charlesworth, B. *Genetics* **137**, 815–827 (1994).
- Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. *Nature* **384**, 346–349 (1996).

Casein kinase I transduces Wnt signals

John M. Peters*†, Renée M. McKay*†, James P. McKay & Jonathan M. Graff*

* Center for Developmental Biology, Department of Molecular Biology, UT Southwestern Medical Center, 5323 Harry Hines Blvd, NB 5.208, Dallas, Texas 75235-9133, USA

† These authors contributed equally to this work

The Wnt signalling cascade is essential for the development of both invertebrates and vertebrates, and is altered during tumorigenesis. Although a general framework for Wnt signalling has been elucidated, not all of the components have been identified. Here we describe a serine kinase, casein kinase I (CKI), which was isolated by expression cloning in *Xenopus* embryos. CKI reproduces several properties of Wnt signals, including generation of complete dorsal axes, stabilization of β -catenin and induction of genes that are direct targets of Wnt signals. Dominant-negative forms of CKI and a pharmacological blocker of CKI inhibited Wnt signals in *Xenopus*. Inhibiting CKI in *Caenorhabditis elegans* generated worms with a *mom* phenotype, indicative of a loss of Wnt signals. In addition, CKI bound to and increased the phosphorylation of dishevelled, a known component of the Wnt pathway. These data indicate that CKI may be a conserved component of the Wnt pathway.

Wnt signalling controls a variety of processes including cell growth, oncogenesis and development^{1,2}. Components of the Wnt cascade are altered in breast cancers in mice and in colon cancers and melanomas in humans^{3,4}. From nematodes to mammals, Wnt signals are key regulators of development^{1,2}; therefore, an understanding of the molecular mechanisms of Wnt signalling has broad implications. In *Xenopus laevis*, Wnt signalling controls formation of the dorsal axis^{5–10}. In *C. elegans*, Wnt signalling controls embryonic polarity, and loss of function of components of the Wnt cascade produces the *mom* (more mesoderm) phenotype in which endoderm is eliminated and replaced by other tissues such as excess pharyngeal muscle^{11,12}.

Through developmental, genetic and biochemical studies, a preliminary outline of Wnt signalling has been proposed^{1,2}. These data were generated in a diverse array of organisms and have converged on a common set of molecules, indicating that the pathway has been conserved. Secreted Wnt ligands form a complex with receptors of the Frizzled family^{13,14}. The signal is then transduced through dishevelled to inhibit glycogen synthase kinase-3 (GSK-3), a negative regulator of the Wnt pathway. The mechanism by which the signal is propagated from dishevelled to GSK-3 is not understood, nor is it known how many steps are involved in this connection^{1,2}. In the absence of Wnt signalling, GSK-3 phosphorylates β -catenin, targeting it for degradation¹⁵. The Wnt-dependent inhibition of GSK-3 alleviates this negative regulation and increases β -catenin levels. β -catenin then binds to the Lef/XTcf family of transcription factors, and the complex translocates to the nucleus, binds directly to the promoters of Wnt-target genes, such as *Xnr3* and *Siamois*^{9,10,16–18}, and alters gene expression.

We isolated a *Xenopus* CKI because it generated a Wnt-like effect in frog embryos. CKI was among the first serine/threonine kinases to be identified and constitutes a family of protein kinases^{19,20}. All CKI isoforms contain highly conserved kinase domains and appear to have similar or identical substrate specificity^{19–21}. Although the biochemistry of these enzymes has been studied in some detail, their biological role and regulation remain poorly defined. Proposed roles include regulation of DNA metabolism, control of protein localization and functions in circadian rhythms^{22–24}. Here we provide gain-of-function, loss-of-function and biochemical evidence that are consistent with the idea that CKI transduces Wnt signals.

CKI generates second dorsal axes

We generated an expression library from two-cell *Xenopus* embryos to identify regulators of early development. Messenger RNA derived from library pools was injected into embryos and a clone from pool 450 produced ectopic dorsal axes. Sib-selection and sequence

analysis identified the clone as a *Xenopus* CKI with the carboxy-terminal tail characteristic of the epsilon isoform. *Xenopus* and human CKI ϵ are greater than 97% identical.

Ectopic Wnt signals induce a complete second axis, mirroring the endogenous role of maternal Wnt signalling in dorsal axis formation^{2,6,25}. CKI was isolated from a maternal source and is therefore expressed at the appropriate time to be involved in generation of the dorsal axis. Although other molecules not in the Wnt pathway, such as chordin and activin, can induce second axes, these axes are typically incomplete and do not include the most anterior structures such as eyes^{26,27}. To characterize the axis-inducing activity of CKI, we injected CKI, a kinase-inactive CKI (CKI (K $\rightarrow$ R))^{20,28}, and a Wnt ligand (Xwnt-8) for comparison. Roughly 80% of CKI (234 out of 298) or Xwnt-8 (58 out of 73) injected embryos formed second dorsal axes, whereas the K $\rightarrow$ R mutant did not, confirming the specificity of the effect (Fig. 1). Notably, CKI and Xwnt-8 produced, at roughly the same frequency (11% and 16%, respectively), embryos in which the second axis was indistinguishable from the primary axis. The alpha isoform of CKI also induced secondary axes (not shown).

Ultraviolet irradiation of *Xenopus* embryos eliminates dorsal axis formation²⁹. Wnt signalling molecules can restore all the dorsal tissues to ultraviolet embryos⁶. We found that CKI (111 out of 114) and Xwnt-8 (60 out of 61), but not the CKI (K $\rightarrow$ R) mutant, produced dorsal axes in ultraviolet embryos with high frequency. CKI, like Wnt signalling components, completely rescued the dorsal axis, producing embryos that were identical to unirradiated, sibling controls (not shown).

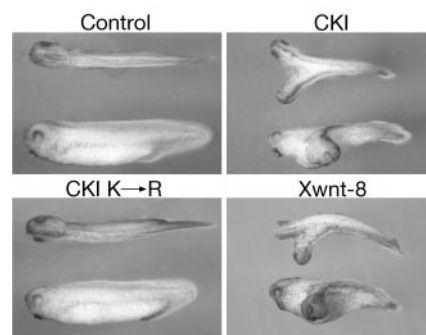


Figure 1 CKI produces complete secondary dorsal axes. Synthetic mRNA encoding CKI (350 pg), mutant CKI (K $\rightarrow$ R) (350 pg) or Xwnt-8 (25 pg) was injected at the eight-cell stage into one ventral vegetal blastomere. Embryos injected with CKI or Xwnt-8, but not inactive CKI (K $\rightarrow$ R), developed complete secondary dorsal axes.

CKI activates Wnt-specific targets

Wnt transduces its signal by post-translationally increasing β -catenin protein levels¹⁵; therefore, stabilization of β -catenin protein is a direct and specific readout of Wnt signalling. To determine whether CKI could increase β -catenin levels, we expressed a Myc-tagged β -catenin alone, with CKI (K $\rightarrow$ R) or with CKI, and analysed the levels of Myc- β -catenin. β -catenin protein levels were analysed before initiation of transcription to ensure that we were assessing a direct effect³⁰. CKI increased β -catenin levels and the inactive K $\rightarrow$ R mutant did not (Fig. 2a); therefore, CKI, like Wnt signals, directly stabilizes β -catenin.

Wnt-stabilized β -catenin binds to the Lef/XTcf family of transcription factors and alters gene expression. As CKI increased β -catenin levels, it seemed likely that it should also induce Wnt specific-markers if it is in the Wnt pathway. Wnt signalling components activate the expression of dorsal, Spemann organizer-specific markers in ventral tissues⁶; in contrast, other signals that can produce second axes or rescue ultraviolet embryos do not induce organizer markers³¹. To test whether CKI could also induce dorsal organizer markers in ventral tissue, two-cell embryos were injected with CKI,

CKI (K $\rightarrow$ R) or Xwnt-8, and ventral marginal zones were excised and analysed for expression of organizer-specific genes as described^{31,32}. CKI and Xwnt-8, but not CKI (K $\rightarrow$ R), induced all the organizer markers (Fig. 2b), including the expression of Siamois and Xnr3, which are direct targets of Wnt signalling^{16,17}.

We used the animal cap assay to extend the molecular comparison between CKI and Wnt signalling. In animal caps, Wnt signals induce the expression of a unique subset of organizer genes, including chordin, Siamois and Xnr3, in the absence of other organizer markers, such as goosecoid or the general mesodermal marker brachyury³¹. We tested whether CKI could activate the expression of the Wnt-specific subset of genes by microinjecting mRNA encoding CKI, CKI (K $\rightarrow$ R), or Xwnt-8 into one-cell embryos, explanting animal caps and analysing gene expression. CKI and Xwnt-8 induced the expression of the same subset of genes

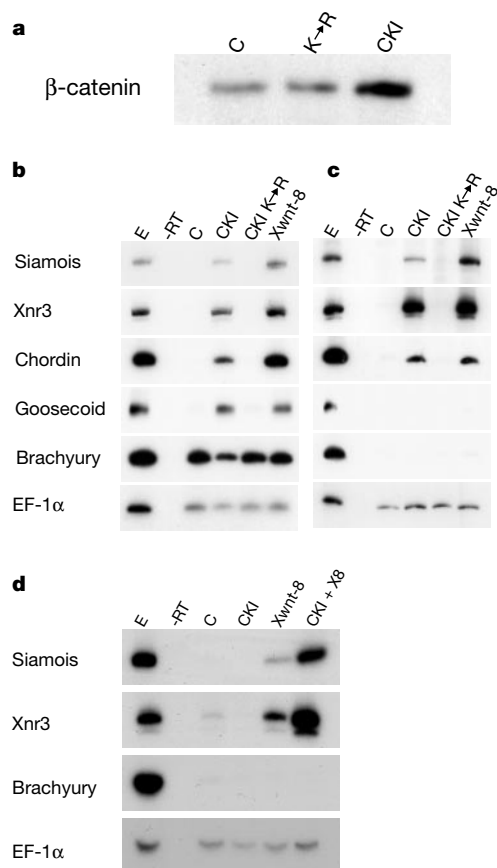


Figure 2 CKI stabilizes β -catenin and induces the expression of Wnt-specific markers. **a**, CKI stabilized β -catenin. One-cell embryos were injected with Myc-tagged β -catenin (100 pg) alone or with either CKI (K $\rightarrow$ R) (350 pg) or CKI (350 pg). Three to four hours after fertilization, embryo lysates were analysed by western blot with an antibody against the Myc-tag. **b**, CKI and Xwnt-8 induced organizer markers in ventral marginal zones (VMZ). VMZs expressing CKI (350 pg), CKI (K $\rightarrow$ R) (350 pg), or Xwnt-8 (25 pg) were analysed by RT-PCR^{31,32}. **c**, CKI induced Wnt-specific markers in the animal cap. Animal caps expressing CKI (350 pg), CKI (K $\rightarrow$ R) (350 pg), or Xwnt-8 (25 pg) were analysed by RT-PCR³¹. **d**, CKI and Xwnt-8 synergized to induce the expression of markers in the animal cap. CKI (25 pg) and Xwnt-8 (X8, 0.5 pg) were injected alone or together and animal caps were analysed.

Sample	n	mean
Uninjected	21	0.0
Xwnt-8 + β -gal (1 ng)	73	1.5
Xwnt-8 + GSK-3 (0.5 ng)	73	1.1
Xwnt-8 + K $\rightarrow$ R (1 ng)	110	0.9
Xwnt-8 + K $\rightarrow$ R (2 ng)	57	0.6
Uninjected	38	0.0
Xwnt-8 (8 pg)	56	1.5
Xwnt-8 + β -gal (1 ng)	85	1.5
Xwnt-8 + D $\rightarrow$ N (0.5 ng)	56	0.9
Xwnt-8 + D $\rightarrow$ N (1 ng)	43	0.4
Uninjected	73	0.0
CKI (350 pg)	44	1.4
CKI + CKI-7 (2.9 ng)	48	0.5
Xwnt-8 (10 pg)	31	1.5
Xwnt-8 + CKI-7	61	0.3
Dsh (500pg)	23	1.4
Dsh + CKI-7	18	0.4
β -cat (1 ng)	36	1.8
β -cat + CKI-7	36	1.8

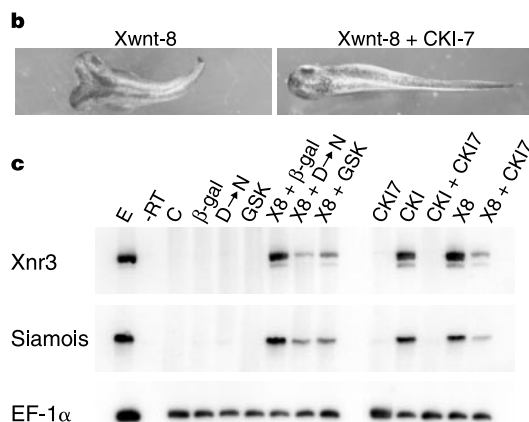


Figure 3 Blocking CKI inhibits Wnt signalling in *Xenopus*. **a**, To evaluate blockade, embryos were scored as no duplication (0), partial duplication (1), or second axis with a head and cement gland (2). For the dominant-negative studies, β -galactosidase (β -gal), GSK-3, CKI (K $\rightarrow$ R) or CKI (D $\rightarrow$ N) were injected at the one-cell stage and then Xwnt-8 was injected at the 16-cell stage into one ventral vegetal blastomere. For the CKI-7 studies, CKI, Xwnt-8, dishevelled (Dsh) and β -catenin (β -cat) with or without CKI-7, were injected into one ventral vegetal blastomere at the eight-cell stage. In part owing to non-specific toxicity, we did not detect a clear and reproducible effect of the inhibitors on primary axis formation. **b**, CKI-7 blocks Wnt function in the second axis assay, performed as described in **a**. **c**, Animal caps expressing β -gal (4 ng), CKI (D $\rightarrow$ N) (4 ng), GSK-3 (1 ng), Xwnt-8 (X8, 10 pg) with β -gal, Xwnt-8 with CKI (D $\rightarrow$ N), Xwnt-8 with GSK-3, CKI-7 (2.9 ng), CKI (350 pg), CKI with CKI-7, or Xwnt-8 with CKI-7 were analysed by RT-PCR.

including the direct targets of Wnt signals, Siamois and Xnr3, whereas the CKI (K → R) mutant did not (Fig. 2c). Thus, in five distinct assays, CKI specifically mimics expression of known Wnt signalling components.

If CKI and Wnt function in a common pathway, then they might interact synergistically. We expressed a relatively low dose of CKI or Xwnt-8 alone and together and assessed the induction of molecular markers. In animal caps, limiting amounts of CKI did not induce expression of Xnr3 and Siamois, whereas Xwnt-8 induced relatively low expression (Fig. 2d). When CKI and Xwnt-8 were co-injected, Siamois and Xnr3 expression were synergistically increased (Fig. 2d). CKI and Xwnt-8 also synergized in rescuing ultraviolet embryos (not shown).

Blocking CKI inhibits Wnt signalling in *Xenopus*

Loss-of-function studies are key to concluding that CKI is in the Wnt pathway; we provide such evidence in two systems, frog and worm (see below). We attempted to derive two dominant-negative forms of CKI that could block Wnt signalling. Kinases that are catalytically inactive owing to K → R mutations are often dominant-

negatives⁷. A D → N mutation in CKI at residue 131 was proposed to be a dominant-negative²⁴. We determined whether these two mutants could block Wnt function in either the double axis or animal cap assay using an approach described for Wnt inhibitors³³ (Fig. 3). Expressing β-galactosidase at the same doses as the mutants did not block Wnt signalling, whereas expression of GSK-3, a negative regulator of Wnt signalling, did (in 3 out of 7 second axis experiments) (Fig. 3). CKI (K → R) inhibited Wnt-dependent second axis formation (3 out of 5) in a dose-dependent manner (Fig. 3a). CKI (D → N) inhibited Wnt in the double axis assay (Fig. 3a); however, CKI (D → N) produced substantial toxicity to embryos which decreased the reliability of the assay. In the animal cap assay, CKI (D → N) was not toxic and blocked Wnt signalling in 100% of experiments (5 out of 5) (Fig. 3c).

Pharmacological inhibitors provide a distinct and alternative approach to blocking CKI activity. CKI-7 is a specific and selective inhibitor of CKI (ref. 34) and, in two assays, was a more potent blocker than either of the dominant-negative forms (Fig. 3). In both the second axis (6 out of 6) and the animal cap assay, CKI-7 inhibited CKI and Wnt in a dose-dependent fashion, at doses that

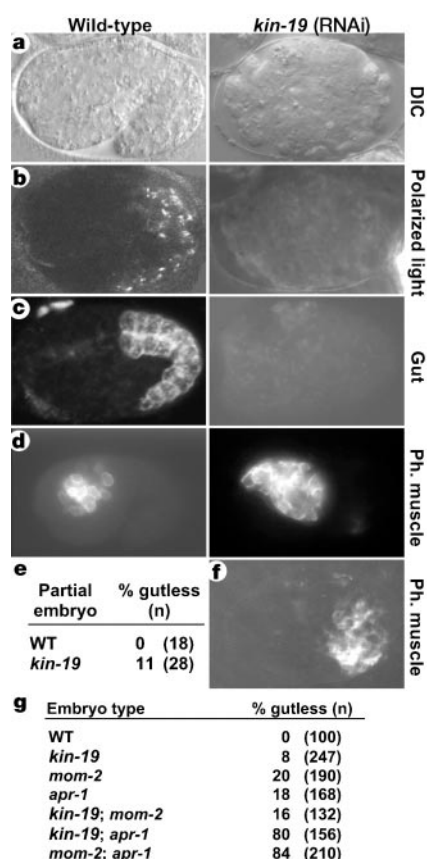


Figure 4 CKI RNA interference produces the *mom* phenotype in *C. elegans*. **a**, DIC micrographs of living embryos. In contrast to a wild-type embryo, *kin-19* RNAi blocked morphogenesis and altered differentiation. **b**, Polarized light microscopy revealed birefringent gut granules (white) in a wild-type embryo but no gut granules in *kin-19* (RNAi)-embryos. **c**, Gut-specific antibody (ICB4) staining showed that, in contrast to wild-type embryos, *kin-19* (RNAi)-embryos lack endoderm. **d**, Staining with a pharyngeal (Ph.) muscle-specific antibody (3NB12) demonstrated that, compared with wild-type embryos, *kin-19* (RNAi)-embryos had an excess of the tissues that can be derived from MS. **e**, **f**, In *kin-19* (RNAi)-embryos, E adopted an MS-like fate. All blastomeres except E were ablated with a laser microbeam in either wild-type (WT) or *kin-19* (RNAi) embryos. **e**, Unlike wild-type, the E-derived tissues from *kin-19* (RNAi) embryos lacked gut granules. **f**, 3NB12-staining showed that E from *kin-19* (RNAi) embryos produced pharyngeal muscle, much of which comes from MS. **g**, *kin-19* (RNAi) was synergistic with *apr-1* (RNAi) but not with *mom-2* (RNAi).

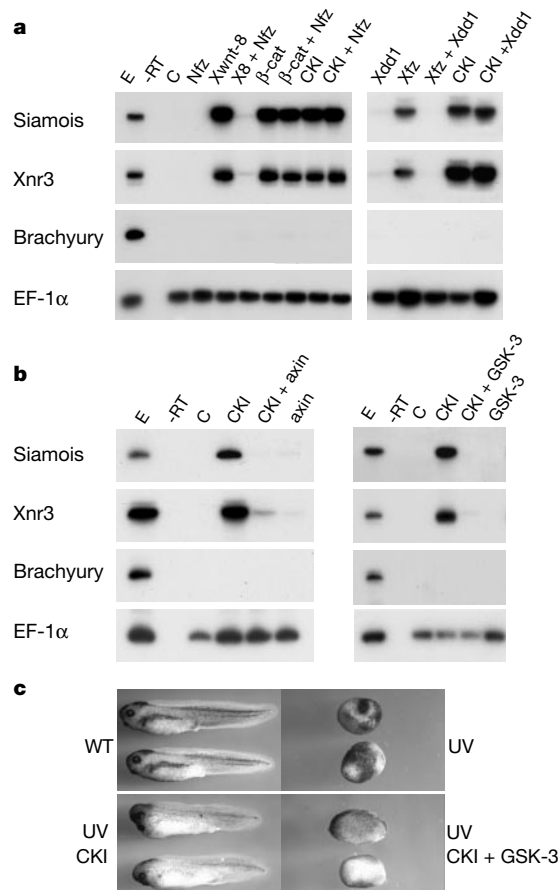


Figure 5 CKI functions between dishevelled and GSK-3. **a**, CKI activity is not blocked by dominant-negative Frizzled (Nfz) or dominant-negative dishevelled (Xdd1). Nfz (1 ng), Xwnt-8 (25 pg), Xwnt-8 (X8) mixed with Nfz, β-catenin (β-cat, 1 ng), β-catenin mixed with Nfz, CKI (350 pg), or CKI mixed with Xdd1 were injected and animal caps analysed for the expression of markers. **b**, GSK-3 and axin blocked CKI in the animal cap assay. CKI (350 pg), axin (1 ng), axin mixed with CKI, GSK-3 (1 ng) and CKI mixed with GSK-3 were injected and animal caps analysed. **c**, GSK-3 blocked the ability of CKI to rescue ultraviolet (UV) embryos. Embryos were UV irradiated and injected with CKI (350 pg) alone or with GSK-3 (1 ng).

approximate the K_i for inhibition of CKI (ref. 34) (Fig. 3). CKI-7 also blocked dishevelled (Fig. 3a); however, CKI-7 did not block β -catenin, showing the specificity of the blockade (Fig. 3a). Overall, these data support the idea that CKI is in the Wnt pathway.

Blocking CKI inhibits Wnt signalling in *C. elegans*

As Wnt signalling is conserved from nematodes to mammals, the function of CKI in the Wnt pathway might also be conserved. In *C. elegans*, Wnt signalling controls the fates of the two daughter cells, MS and E, produced by the EMS blastomere^{11,12}. MS produces mesoderm, and E produces only endoderm. Active Wnt signalling is essential for specification of the E blastomere, and in the absence of Wnt signals E is transformed into MS. The resultant phenotype, *mom*, is a worm without endoderm and with excess muscle^{11,12}.

Database analysis identified two *C. elegans* genes, CO3C10.1 (*kin-19*) and F46F2.2 (*kin-20*), with high, ~50%, sequence identity to *Xenopus* CKI. In *C. elegans*, several components of the Wnt pathway were characterized by inhibiting their function through RNA interference (RNAi)^{12,35–37}. This approach resulted in a *mom* phenotype of variable penetrance, ranging from 2% (*mom-5*, Frizzled) to 100% (*wrm-1*, β -catenin). RNAi with *kin-20* produced a sterile phenotype, precluding the ability to score embryonic phenotypes

(not shown). When RNAi was performed with *kin-19*, 8% of the resulting worm embryos had the *mom* phenotype as judged by a variety of criteria (Fig. 4). First, differential interference contrast (DIC) microscopy demonstrated loss of both morphogenesis and endodermal differentiation with excess mesoderm formation (Fig. 4a). Second, polarized light microscopy revealed that the embryos contained no gut granules, consistent with a loss of endodermal fates (Fig. 4b). Third, the absence of endoderm was confirmed by tissue-specific antibody staining (Fig. 4c). Fourth, tissue-specific antibody staining showed that these embryos had excess pharyngeal muscle, an MS derivative (Fig. 4d). These features are all hallmarks of the *mom* phenotype, which results from loss of Wnt signals.

To determine whether this phenotype was due to an E-to-MS transformation, we used RNAi followed by laser ablation of all blastomeres except E and analysed the descendants of E^{11,12}. When wild-type embryos were ablated and analysed, E always produced endoderm (Fig. 4e). In roughly 11% of *kin-19* RNAi-treated embryos, E did not form endoderm (Fig. 4e). Instead, these partial embryos generated pharyngeal muscle (Fig. 4f). In addition, we undertook RNAi with just the C-terminal domain of *kin-19*, a region outside the kinase domain which does not share homology with other protein kinases. This determined whether the *mom* phenotype might result from promiscuous inhibition of another kinase; if so, then RNAi with the tail alone would not produce the *mom* phenotype. As with the full-length clone, however, RNAi with the tail of *kin-19* produced embryos that lacked endoderm and contained excess pharyngeal muscle (not shown). As another specificity control, we inhibited casein kinase II (CKII), another serine kinase that binds to dishevelled³⁸; the resultant CKII RNAi embryos did not have the *mom* phenotype (not shown). Thus, RNAi of CKI in the worm specifically produced the same phenotype, *mom*, as loss of Wnt signalling.

Some components of the Wnt cascade, including the receptor, GSK-3 and β -catenin, have a linear relationship downstream of Wnt and are directly modulated by the ligand¹². Other components, such as adenomatous polyposis coli (APC) modify the cascade but are not directly downstream of Wnt¹². When two components of the linear pathway are simultaneously eliminated, no synergy of the *mom* phenotype is observed¹². In contrast, when a linear component and APC are eliminated together, there is a marked increase in the percentage of *mom* mutant embryos¹² (Fig. 4g). To determine any synergy with CKI and other components of the Wnt cascade, we simultaneously decreased the function of CKI and either MOM-2 (Wnt) or APR-1 (APC) by RNAi and analysed the resultant embryos. RNAi combining *kin-19* and *mom-2* had no effect on the penetrance of the phenotype (Fig. 4g). In contrast, we observed a synergistic enhancement of the mutant phenotype by inhibiting both *kin-19* and *apr-1* (Fig. 4g). This was observed with other linear components of the Wnt pathway and is consistent with CKI functioning directly in the cascade¹².

CKI functions between dishevelled and GSK-3

We used epistasis tests to determine where CKI functions in the Wnt pathway. Initially, we analysed Wnt, dishevelled or β -catenin function while blocking CKI and found that CKI functions downstream of Wnt and dishevelled and upstream of β -catenin (Fig. 3). As a complementary approach, we assessed CKI activity in the presence of Wnt pathway inhibitors. In the animal cap assay, Nfz^{14,39}, a dominant-negative Frizzled, and Xdd1⁴⁰, a dominant-negative dishevelled, both blocked Wnt, but neither blocked CKI (Fig. 5a). GSK-3 and axin function downstream of Frizzled and dishevelled and both are negative regulators of the Wnt pathway^{14,41}. Both GSK-3 and axin inhibited CKI in the animal cap assay (Fig. 5b). GSK-3 (Fig. 5c) and axin (not shown) also blocked CKI function in the ultraviolet rescue assay. Thus, CKI functions downstream of dishevelled and upstream of GSK-3.

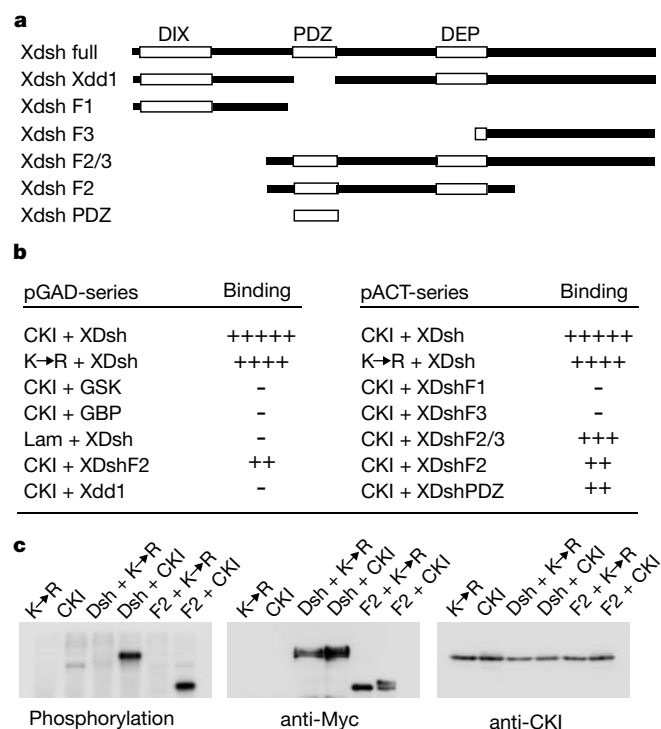


Figure 6 CKI binds to and increases the phosphorylation of dishevelled. **a**, The dishevelled protein (Xdsh) and fragments used in the yeast two-hybrid and phosphorylation assays. **b**, CKI bound dishevelled in a yeast two-hybrid assay. CKI and CKI (K → R) baits interacted strongly with Xdsh. CKI also interacted with Xdsh fragments 2/3 (F2/3; amino acids 231–737), fragment 2 (231–531), and the PDZ domain (301–381). CKI did not interact with GSK-3, GBP, Xdsh fragment 1 (1–289), Xdsh fragment 3 (485–737) or a dominant negative form of Xdsh, Xdd1, that lacks the PDZ domain. Yeast two-hybrid was done by standard methods^{42,50}. Relative strength of binding was determined by a liquid β -galactosidase assay. To establish reproducibility, two different reporter vectors, pGAD and pACT (Clontech), were used. None of the clones activated in the absence of the interacting partner. **c**, *Xenopus* oocytes were injected with mRNA encoding either CKI or CKI (K → R) (15 ng) in the presence of either Myc-tagged dishevelled (Dsh, 30 ng) or Myc-tagged dishevelled fragment 2 (F2, 30 ng), incubated with ³²P_i, lysed and immunoprecipitated via the Myc-tag. Protein levels were analysed by western blots against the Myc-epitope or CKI.

CKI and dishevelled interact

Another approach to positioning CKI in the Wnt pathway is to determine whether CKI binds to or phosphorylates components of the Wnt cascade. We employed yeast two-hybrid with CKI in both an unbiased screen and a directed approach to evaluate binding. We screened 1×10^5 clones of a mouse neonatal brain library and, out of five positives, one was mouse dishevelled. We then determined whether CKI would also bind *Xenopus* dishevelled and, as the epistasis studies placed CKI between dishevelled and GSK-3, we also tested interaction with the other components of the Wnt pathway, GSK-3 and GBP⁴², that are consistent with this position^{1,2}. Both CKI and CKI (K → R) bound to dishevelled but not to the other molecules tested, showing the specificity of the interaction (Fig. 6b). The catalytic domain of CKI also interacted with dishevelled (not shown).

Presumably, CKI regulates Wnt signalling by phosphorylating an element of the cascade. Thus we expressed, in *Xenopus* oocytes, Myc-tagged components of the Wnt pathway with or without CKI or CKI (K → R) and incubated the oocytes in ³²P_i. Next, we immunoprecipitated the epitope-tagged constructs and evaluated phosphorylation levels. CKI did not increase the phosphorylation of GSK-3 or β-catenin (not shown); however, it did increase the phosphorylation of dishevelled, whereas CKI (K → R) did not, confirming specificity (Fig. 6c).

To define the region where CKI binds, we divided dishevelled into three roughly equal fragments and tested these fragments for interaction in yeast. Unlike the first and third fragments, the middle fragment of dishevelled bound to CKI (Fig. 6b). Consistent with the two-hybrid data, CKI increased the phosphorylation of the middle fragment (Fig. 6c). This middle region of dishevelled contains an 80-amino-acid PDZ protein-interaction domain that is critical for wild-type function⁴⁰. Notably, CKI directly bound this domain (Fig. 6b). Deleting these 80 amino acids converts dishevelled into a dominant-negative form, Xdd1 (ref. 40), which did not interact with CKI (Fig. 6b).

Discussion

On the basis of a concordance of data obtained in *Xenopus* and *C. elegans*, we propose that CKI is a conserved component of the Wnt signalling cascade. The confluence of results—including a demonstration of specificity in each system, synergy in both systems, distinct and complementary approaches (gain- and loss-of-function), the use of two divergent organisms and the different roles of Wnt signalling in the two systems (endoderm versus dorsal axis)—markedly strengthens this conclusion. Furthermore, biochemical studies show that CKI interacts with and increases the phosphorylation of dishevelled, a known component of the Wnt pathway.

At what position in the Wnt pathway does CKI function? CKI and APR synergize in producing the *mom* phenotype which suggests that CKI has a linear relationship with Wnt, the Wnt receptor (Frizzled), GSK-3 and β-catenin in the cascade¹². Furthermore, epistasis studies in the frog indicate that CKI is downstream of dishevelled and upstream of GSK-3, which is consistent with CKI increasing β-catenin protein levels. In addition, CKI directly binds dishevelled, a component of the Wnt pathway that is upstream of GSK-3 (ref. 2). Together, these data indicate that CKI may function between dishevelled and GSK-3.

Dishevelled phosphorylation increases in response to Wnt signalling⁴³. Possibly, CKI is the responsible kinase, as it binds to and increases the phosphorylation of dishevelled. Xdd1, a form of dishevelled with an 80-amino-acid deletion, functions as a dominant-negative⁴⁰. These 80 amino acids, when expressed in isolation, bind to CKI whereas Xdd1 does not; thus, these 80 amino acids may be critical for the CKI–dishevelled interaction. The abrogation of CKI binding may account for the dominant-negative action of Xdd1.

The epistasis studies suggest that CKI is downstream of dishevelled,

yet CKI phosphorylates dishevelled. Perhaps the CKI-dependent phosphorylation of dishevelled is not responsible for the biological role that CKI plays in the Wnt pathway. For example, the phosphorylation of dishevelled may just be a marker of CKI activation. Equally plausible, however, is that the CKI-dependent phosphorylation of dishevelled modulates or modifies downstream signalling. Further biochemical studies such as phosphorylation-site mapping and mutagenesis will be required to determine the mechanism of action.

Wnt signalling is important in development and carcinogenesis^{1,2}. A detailed biochemical understanding of the Wnt cascade may allow direct manipulation of those processes. Given the extensive characterization of protein kinases, the discovery that CKI is in the Wnt pathway should promote a better characterization of the mechanisms of Wnt signalling and of the Wnt-dependent regulation of cellular events. □

Methods

Library generation and sib selection

Poly(A)⁺ RNA from 2,000 two-cell stage embryos was synthesized into complementary DNA with the Superscript Plasmid System (GibcoBRL), ligated into *SalI*–*NotI*-digested pCS105, transformed and plated. Pooled plasmid DNA was isolated and linearized with *AscI*, and mRNA was synthesized as described⁶. Single clones from pool 450 were isolated by sib-selection roughly as described⁶.

Microinjection, dissection and RT-PCR analysis

mRNA was synthesized for microinjections as described³¹. Animal cap, marginal zone, double axis, and ultraviolet rescue microinjection assays were performed as described^{31,32,44}. RT-PCR analyses were performed as described^{31,32,44}.

Kinase mutant CKI

To inactivate CKI, residue 38 was changed from lysine to arginine (K → R)^{20,28} or residue 131 was changed from aspartate to asparagine (D → N)²⁴ by PCR⁴⁵.

C. elegans

Standard methods were used to culture *C. elegans* wild-type strain Bristol N2. Antibody staining with pharyngeal muscle-specific monoclonal antibody 3NB12 (ref. 12) and gut-specific monoclonal antibody ICB4 (ref. 46) was done as described⁴⁷, except that monoclonal cell supernatants were diluted 1:20. Templates for RNA were full-length cDNAs from C03C10.1 (*kin-19*) and F46F2.2 (*kin-20*) generated by RT-PCR. Partial cDNAs for *mom-2* and *apr-1* were generated in a similar fashion. RNA was generated with Megascript (Ambion). The sense and antisense RNAs were annealed for 15 min at 68 °C in TE and then for 1 h at 37 °C. The annealed RNA (1 mg ml^{−1}) was injected into both gonad arms of young adult hermaphrodites. Embryos for analysis were collected 3–24 h after injection. Laser ablations were as described^{1,48}.

Phosphorylation

Xenopus oocytes were defolliculated, injected and then cultured for 24 h at 18 °C with 0.6 μCi ³²P_i μl^{−1}. Samples were lysed, centrifuged at 10,000g for 5 min, and the supernatants immunoprecipitated and subjected to SDS–PAGE and autoradiography or western blotting (anti-Myc, 9E10; anti-CKI antibodies C40520, Transduction Laboratories).

Received 13 July; accepted 19 August 1999.

- Cadigan, K. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
- Moon, R., Brown, J. & Torres, M. WNT's modulate cell fate and behavior during vertebrate development. *Trends Genet.* **13**, 157–162 (1997).
- Nusse, R. & Varmus, H. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99–109 (1982).
- Rubinfeld, B. *et al.* Stabilization of β-catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
- Heasman, J. *et al.* Overexpression of cadherins and underexpression of β-catenin inhibit dorsal mesoderm induction in early *Xenopus* embryos. *Cell* **79**, 791–803 (1994).
- Smith, W. & Harland, R. Injected Xwnt-8 RNA acts early in *Xenopus* embryos to promote formation of a vegetal dorsalizing center. *Cell* **67**, 753–765 (1991).
- He, X., Saint-Jeannet, J., Woodgett, J., Varmus, H. & Dawid, I. Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos. *Nature* **374**, 617–622 (1995).
- Pierce, S. & Kimelman, D. Regulation of Spemann organizer formation by the intracellular kinase Xgsk-3. *Development* **121**, 755–765 (1995).
- Molenaar, M. *et al.* XTCF-3 transcription factor mediates β-catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Behrens, J. *et al.* Functional interactions of β-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
- Thorpe, C., Schlesinger, A., Carter, J. & Bowerman, B. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).

12. Rocheleau, C. *et al.* Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
13. Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a wingless receptor. *Nature* **382**, 225–230 (1996).
14. Xu, Q., D'Amore, P. & Sokol, S. Functional and biochemical interactions of Wnts with FrzA, a secreted Wnt antagonist. *Development* **125**, 4767–4776 (1998).
15. Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β -catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
16. McKendry, R., Hsu, S., Harland, R. & Grosschedl, R. LEF-1/TCF proteins mediate wnt-inducible transcription from the *Xenopus* nodal-related 3 promoter. *Dev. Biol.* **192**, 420–431 (1997).
17. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. & Kimelman, D. A β -catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
18. Carnac, G., Kodjabachian, L., Gurdon, J. & Lemaire, P. The homeobox gene Siamois is a target of the Wnt dorsalization pathway and triggers organiser activity in the absence of mesoderm. *Development* **122**, 3055–3065 (1996).
19. Tuazon, P. & Traugh, J. in *Advances in Second Messenger and Phosphoprotein Research* Vol. 23, 123–164 (Raven, New York, 1991).
20. Fish, K., Cegijska, A., Getman, M., Landes, G. & Virshup, D. Isolation and characterization of human casein kinase I epsilon (CKI), a novel member of the CKI gene family. *J. Biol. Chem.* **270**, 14875–14833 (1995).
21. Songyang, Z. *et al.* A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol. Cell. Biol.* **16**, 6486–6493 (1996).
22. Hoekstra, M. *et al.* HRR25, a putative protein kinase from budding yeast: association with repair of damaged DNA. *Science* **253**, 1031–1034 (1991).
23. Kloss, B. *et al.* The *Drosophila* clock gene double-time encodes a protein closely related to human casein kinase I ϵ . *Cell* **94**, 97–107 (1998).
24. Zhu, J. *et al.* Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEKK1. *Cell* **93**, 851–861 (1998).
25. McMahon, A. & Moon, R. Ectopic expression of the proto-oncogene int-1 in *Xenopus* embryos leads to duplication of the embryonic axis. *Cell* **58**, 1075–1084 (1989).
26. Thomsen, G. *et al.* Activins are expressed early in *Xenopus* embryogenesis and can induce axial mesoderm and anterior structures. *Cell* **63**, 485–493 (1990).
27. Sasai, Y. *et al.* *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779–790 (1994).
28. DeMaggio, A., Lindberg, R., Hunter, T. & Hoekstra, M. The budding yeast HRR25 gene product is a casein kinase I isoform. *Proc. Natl Acad. Sci.* **89**, 7008–7012 (1992).
29. Scharf, S. & Gerhart, J. Axis determination in eggs of *Xenopus laevis*: A critical period before first cleavage, identified by the common effects of cold, pressure, and ultraviolet irradiation. *Dev. Biol.* **99**, 75–87 (1983).
30. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696 (1982).
31. LeSueur, J. & Graff, J. Spemann organizer activity of Smad10. *Development* **126**, 137–146 (1999).
32. Graff, J., Thies, R., Song, J., Celeste, A. & Melton, D. Studies with a *Xenopus* BMP receptor suggest that ventral mesoderm-inducing signals override dorsal signals *in vivo*. *Cell* **79**, 169–179 (1994).
33. Hoppler, S., Brown, J. & Moon, R. Expression of a dominant-negative Wnt blocks induction of MyoD in *Xenopus* embryos. *Genes Dev.* **10**, 2805–2817 (1996).
34. Chijiwa, T., Hagiwara, M. & Hidaka, H. A newly synthesized selective casein kinase I inhibitor, N-(2-aminoethyl)-5-chloroisouquinoline-8-sulfonamide, and affinity purification of casein kinase I from bovine testis. *J. Biol. Chem.* **264**, 4924–4927 (1989).
35. Tabara, H., Grishok, A. & Mello, C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430–431 (1998).
36. Sharp, P. RNAi and double-strand RNA. *Genes Dev.* **13**, 139–141 (1999).
37. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
38. Willert, K., Brink, M., Wodarz, A., Varmus, H. & Nusse, R. Cusein kinase 2 associates with and phosphorylates Dishevelled. *EMBO J.* **16**, 3089–3096 (1997).
39. Deardorff, M., Tan, C., Conrad, L. & Klein, P. Frizzled-8 is expressed in the Spemann organizer and plays a role in early morphogenesis. *Development* **125**, 2687–2700 (1998).
40. Sokol, S. Analysis of Dishevelled signaling pathways during *Xenopus* development. *Curr. Biol.* **6**, 1456–1467 (1996).
41. Zeng, L. *et al.* The mouse fused locus encodes axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. *Cell* **90**, 181–192 (1997).
42. Yost, C. *et al.* GSK-3, an inhibitor of *Xenopus* development and oncogenesis. *Cell* **93**, 1031–1041 (1998).
43. Yanagawa, S., van Leeuwen, F., Wodarz, A., Klingensmith, J. & Nusse, R. The dishevelled protein is modified by Wingless signaling in *Drosophila*. *Genes Dev.* **9**, 1087–1097 (1995).
44. Graff, J., Bansal, A. & Melton, D. *Xenopus* Mad proteins transduce distinct subsets of signals for the TGF β superfamily. *Cell* **85**, 479–487 (1996).
45. Ho, S., Hunt, H., Horton, R., Pullen, J. & Pease, L. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
46. Bowerman, B., Draper, B., Mello, C. & Priess, J. Cell interactions involved in development of the bilaterally symmetrical intestinal valve cells during embryogenesis in *Caenorhabditis elegans*. *Development* **116**, 1113–1122 (1992).
47. Bowerman, B., Draper, B., Mello, C. & Priess, J. The material gene *skin-1* encodes a protein that is distributed unequally in early *C. elegans* embryos. *Cell* **74**, 443–452 (1993).
48. Epstein, H. & Shakes, D. in *Methods in Cell Biology* (eds Wilson, L. & Matsudaira, P.) (Academic, San Diego, 1995).
49. Krieg, P. A., Varnum, S., Wormington, M. & Melton, D. A. The mRNA encoding elongation factor 1 α (EF-1 α) is a major transcript at the mid blastula transition in *Xenopus*. *Dev. Biol.* **133**, 93–100 (1989).
50. Chien, C., Bartel, P., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl Acad. Sci.* **88**, 9578–9582 (1991).

Acknowledgements

We thank L. Avery, M. Henkemeyer, J. Jiang, R. Lalan, L. Meng, L. Parada, and members of the Graff lab for support and comments. We also thank X. Cao, M. Cobb, F. Constantini (axin), C. Cowan, R. Harland (Xwnt-8, β -galactosidase, pCS105), J. Heasman, D. Kessler, D. Kimelman (GBP, GSK-3), P. Klein (XFz8, Nfz), R. Lin, R. Moon, B. Powell (mouse neonatal brain yeast two-hybrid library), J. Priess (3NB12 and ICB4 antibodies), S. Sokol (Xdsh, Xdd1), D. Turner (pCS2 + MT), and C. Wylie (β -catenin) for providing reagents and technical advice. J.P.M. was supported by NIH and Established Investigator AHA awards to L. Avery, in whose laboratory the *C. elegans* experiments were conducted. This work was supported by an NIH award to J.M.G. J.M.G. is a March of Dimes Basil O'Connor Scholar and a Charles E. Culpeper Medical Scholar.

Correspondence and requests for materials should be addressed to J.M.G. (e-mail: graff02@uts.wmed.edu). The accession number for *Xenopus* CKI epsilon is: AF183394

A distance to the galaxy NGC4258 from observations of Cepheid variable stars

Eyal Maoz^{*†}, Jeffrey A. Newman[†], Laura Ferrarese[‡], Peter B. Stetson[§], Stephen E. Zepf^{||}, Marc Davis[†], Wendy L. Freedman[¶] & Barry F. Madore[¶]

^{*} NASA Ames Research Center, MS 245-3, Moffett Field, California 94035-1000, USA

[†] Department of Astronomy, University of California, Berkeley, California 94720, USA

[‡] California Institute of Technology, MS 105-24, Pasadena, California 91125, USA

[§] Dominion Astrophysical Observatory, 5071 W. Saanich Road, Victoria, British Columbia, Canada V8X 4M6

^{||} Department of Astronomy, PO Box 208101, Yale University, New Haven, Connecticut 06520, USA

[¶] Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA

[#] NASA/IPAC Extragalactic Database, Infrared Processing and Analysis Center, Jet Propulsion Laboratory, California Institute of Technology, MS 100-22, Pasadena, California 91125, USA

Cepheid variable stars pulsate in a way that is correlated with their intrinsic luminosity, making them useful as 'standard candles' for determining distances to galaxies; the potential systematic uncertainties in the resulting distances have been estimated¹ to be only 8–10%. They have played a crucial role in establishing the extragalactic distance scale and hence the value of the Hubble constant. Here we report observations of Cepheids in the nearby galaxy NGC4258; the distance calculated from the Cepheids is 8.1 ± 0.4 Mpc, where the uncertainty does not include possible systematic errors. There is an independently determined geometric distance to this galaxy of 7.2 ± 0.5 Mpc, based on the observed proper motions of water masers orbiting the central black hole²; the distances differ by 1.3σ . If the maser-based distance is adopted and the Cepheid distance scale revised accordingly, the derived value of the Hubble constant would increase by $12 \pm 9\%$, while the expansion age of the Universe would decrease by the same amount.

Over the past 20 years, substantial improvements have been made in the Cepheid distance scale^{3–5}. These stars have remained the primary extragalactic distance indicator since 1929 because of the small observed scatter in the relationship between their pulsation period and luminosity, their large numbers, which allow many independent measures of the distance to a galaxy, and the simplicity of the basic physics underlying their variability. The Cepheid period–luminosity relation derived from observations of stars in the Large Magellanic Cloud (LMC) has formed the basis for the calibration of the Hubble Space Telescope (HST) Key Project on the Extragalactic Distance Scale (ref. 6, henceforth the Key Project) and of many other recent extragalactic distance measurements⁷. Cepheid distances to galaxies as far away as 25 Mpc have been reliably measured using the HST with random errors of a few per cent; the results have been used to calibrate a number of secondary distance methods and hence provide estimates of the Hubble constant.

Today, the largest identified potential sources of systematic error in the Cepheid distance scale are the zero point of the Cepheid period–luminosity relation (or alternatively, the adopted distance to the LMC), difficulties in the calibration of the HST Wide Field and Planetary Camera 2 (WFPC2) which has been used for most modern studies, and the possible effects of differences in the chemical composition of stars on Cepheid distance measurements. These error values have been judged to be $\pm 6.5\%$, $\pm 4.5\%$, and $\pm 4\%$ in distance, respectively¹, with substantial non-gaussianity possible; for instance, recent measurements of the distance to the LMC range from ~ 40 to ~ 55 kpc, with a distribution skewed

towards lower values⁸. Given the remaining uncertainties affecting the application of the LMC-calibrated Cepheid period–luminosity relation to galaxies observed with the HST, further tests of the extragalactic distance scale are needed.

The spiral galaxy NGC4258 presents new opportunities to test and potentially to improve the calibration of the Cepheid period–luminosity relation because of the precision with which its distance has been measured, independently of the conventional ladder of astronomical distance scales⁹. This distance (7.2 Mpc) has been determined using its apparently simple, keplerian, circumnuclear disk, which is delineated by line-emitting water masers that orbit a supermassive black hole at its centre^{10,11}. The total estimated uncertainty in this distance is ± 0.3 Mpc if the disk is assumed to be circular; if non-zero eccentricities are allowed, the uncertainty increases to ± 0.5 Mpc (we adopt this value). The direct, geometric methods used are believed to have minimal unknown systematic uncertainties. Combining the observed rotation velocities with the measured centripetal acceleration in the disk¹² or the observed proper motions of the maser sources² allows independent measurements of the physical size of the disk; comparing these to its observed angular extent yields the distance to the galaxy centre using simple geometry. The two methods of calculating this distance (proper motions and accelerations) yield results in agreement with each other to 1%.

We have therefore observed a portion of NGC4258 to search for Cepheids using the WFPC2 camera and the Hubble Space Telescope on 11 epochs in 1998 with both the F555W and F814W filters (similar to standard V and I filters). Details of the observations and the analysis procedures used are described and full photometric

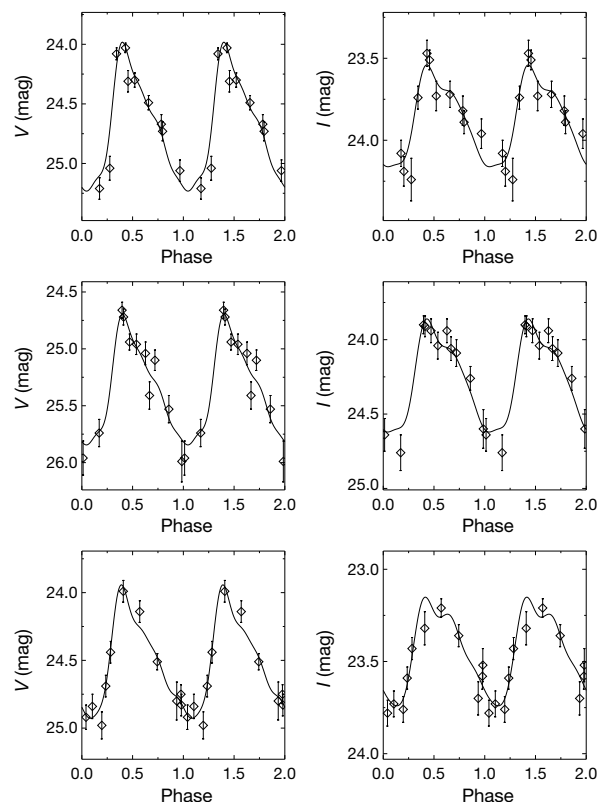


Figure 1 Light curves for three representative examples of the Cepheids in NGC4258. Shown are the measured magnitudes of each star (diamond symbols) and their uncertainty values (error bars); the template Cepheid light curves (solid lines) are derived from observation of stars in the Large Magellanic Cloud (LMC)³⁰ that have been used to fit the data. The left panel for each star shows the variation of the V magnitude of each star versus the phase of the variation, which makes the Cepheid pulsation pattern clear. Similarly, the right panel plots the I magnitude variation of the Cepheids.

results provided elsewhere (J.A.N. *et al.*, manuscript in preparation). We obtained photometry by fitting point-spread functions to the data using two software packages (DAOPHOT/ALLFRAME¹³ and DoPHOT^{7,14}), following procedures¹⁵ similar to those used for the Key Project, and converted the results to standard *V* and *I* filter photometric systems. Cepheid searches were performed using two different algorithms with the ALLFRAME data set^{17,30}, and a third for the DoPHOT data¹⁵.

We have identified and determined light curves, periods, mean magnitudes, and colours for 15 Cepheids. All of these stars fulfil four criteria: they are identified as variable by all three search techniques, fit a Cepheid template light curve with reasonable χ^2 , visibly vary in blink comparisons in both F555W and F814W images, and have negligible statistical probability of being misidentified non-variables. Light curves illustrating the variability of several of the Cepheids found are plotted in Fig. 1. The DoPHOT NGC4285 period–luminosity relations are plotted in Fig. 2. We adopted the DoPHOT photometry for all the main conclusions reported here, as ALLFRAME photometry yielded distances from Cepheids on the two WFPC2 chips used with an internal discrepancy of 0.24 mag (a 1.8σ difference possibly due to an error in aperture corrections), whereas the corresponding DoPHOT results agreed within 0.07 mag (0.5σ); however, as an additional check we also provide ALLFRAME values.

In Key Project procedures, the differences between the observed and intrinsic colours of Cepheids are used to correct for the effects of extinction by dust, assuming a standard Galactic reddening law¹⁸. We applied this correction star by star and then robustly averaged the resulting distance moduli, as this method gives values for small data sets that best match those that would be obtained using Key Project methods with more Cepheids (J.A.N. *et al.*, manuscript in preparation). This technique yields a distance with a statistical uncertainty (derived from the period–luminosity relation fit) of ± 0.07 mag. We would expect a possible 1σ systematic error of ± 0.04 mag from the uncertainties in photometry because of the difficulty of obtaining accurate aperture corrections in our fields, which contained relatively few bright, isolated stars. This error should be highly correlated between *V* and *I* magnitude measurements and thus not propagated through the reddening correction. However, to be conservative we instead adopt the entire ALLFRAME–DoPHOT distance modulus difference (0.10 mag) as an estimate of possible errors affecting only photometry for NGC4258. Adding this in quadrature to the random error, the total uncertainty unique to our determination of a Cepheid distance to NGC4258 is ± 0.12 mag.

This measurement is subject to a number of potential sources of systematic error that also affect Key Project distance determinations, as described in Table 1; their possible contributions have been estimated to a total ± 0.18 mag (ref. 19). Further description of these sources of error may be found elsewhere²⁰. Correcting for differences in heavy-element content between the field we have studied and the LMC (ref. 21), using the results of previously published studies of NGC4258 (ref. 22), would lead to an increase of 0.08 ± 0.06 mag in our distance modulus value. However, due to the lack of agreement on the magnitude and sign of this effect^{1,21}, the Key Project has refrained from applying such corrections, instead considering the resulting uncertainty to be a possible systematic error. We thus do likewise, and obtain a Cepheid distance modulus to NGC4258 of 29.54 ± 0.12 mag (unique to this determination) ± 0.18 mag (systematic uncertainties in Key Project distances), corresponding to a metric distance of 8.1 ± 0.4 Mpc ± 0.7 Mpc. When treated in the same way, the ALLFRAME results yield a distance modulus of 29.64 ± 0.13 mag, corresponding to a metric distance of 8.5 ± 0.5 Mpc. The distance to NGC4258 derived from observations of Cepheids is thus appreciably greater than the maser distance of 7.2 ± 0.5 Mpc (ref. 2).

The Cepheid and maser distances differ by 1.0σ if we add together

Table 1 Uncertainties in the distance measurement

Symbol	Source	Error
S_1	Systematic errors in LMC <i>P</i> – <i>L</i> calibration	
	LMC true modulus (<i>a</i>)	± 0.13
	LMC <i>P</i> – <i>L</i> zero point (<i>b</i>)	± 0.02
	<i>a</i> and <i>b</i> added in quadrature	± 0.13
S_2	Systematic errors in WFPC2 photometry	
	HST WFPC2 <i>V</i> –band zero point (<i>c</i>)	± 0.03
	HST WFPC2 <i>I</i> –band zero point (<i>d</i>)	± 0.03
	$\sqrt{c^2(1-R)^2 + d^2R^2}$ for $R = A(V)/E(V-I)$	± 0.09
S_3	Average metallicity correction	$+0.08 \pm 0.06$
S_4	Systematic errors unique to NGC4258 photometry (from DoPHOT–ALLFRAME comparison)	± 0.10
R_1	Random error in the NGC4258 extinction-corrected distance modulus	
	NGC4258 <i>P</i> – <i>L</i> fit (<i>V</i> band; <i>g</i>)	± 0.07
	NGC4258 <i>P</i> – <i>L</i> fit (<i>I</i> band; <i>h</i>)	± 0.05
	<i>g</i> and <i>h</i> partially correlated	± 0.07
R_{tot}	Errors only affecting this determination (S_4 and R_1 added in quadrature)	± 0.12
S_{tot}	Systematic errors in techniques used (S_1 , S_2 and S_3 added in quadrature)	± 0.18

All values listed are in magnitudes; for small differences, the corresponding fractional error in distance may be obtained by multiplying by 0.46. For those potential systematic errors which uniformly affect all Cepheid distances obtained in the same manner as ours, we have adopted the uncertainty estimates of the Key Project²⁰. The estimated systematic errors in WFPC2 photometry (S_2) include uncertainties in zero points and the ‘long versus short’ uncertainty, added in quadrature. The metallicity of the field we studied in NGC4258 matches the average metallicity of Key Project galaxies used to calibrate the Tully–Fisher relation, surface brightness fluctuation distances, and the peak luminosities of type Ia supernovae; we therefore treat the uncertainties in metallicity correction as a systematic error common to this distance measurement and those of the Key Project (S_3). The random uncertainty in the Cepheid distance modulus (R_1) has been estimated from the standard error of the extinction-corrected moduli for individual stars. We have adopted a conservative estimate of the total photometric uncertainties unique to NGC4258 (S_4) of 0.10 mag based on the difference between the overall distance moduli they yield. For chip 2, the mean difference between ALLFRAME and DoPHOT magnitudes for 24 bright, isolated stars was 0.026 ± 0.049 (standard deviation) mag for *V*, and 0.088 ± 0.049 mag for *I*. For chip 3, the mean difference for 30 stars was 0.025 ± 0.027 mag for *V*, and 0.088 ± 0.046 mag for *I*. Mean magnitudes for Cepheids yielded results consistent with these to within 1σ , albeit with much larger standard errors. We consider the / results from chip 3 to be an aberration closely related to the discrepant distance moduli obtained with ALLFRAME on the two chips; however, even if it were included, the mean ALLFRAME–DoPHOT magnitude difference among the four cases would be 0.04 mag.

in quadrature our measurement uncertainty of 0.4 Mpc, the Key Project systematic error estimate of 0.7 Mpc, and the maser distance error estimate of 0.5 Mpc; potential systematic errors in either technique do not seem to have been grossly underestimated. However, if we assume that the maser distance is correct (up to its stated uncertainty) and wish to determine whether a revision of the Cepheid distance scale may be desirable, these prior estimates of systematic errors in the Cepheid distance scale are irrelevant to calculations of significance. In that case, there is a 1.3σ (80% significance) discrepancy (1.6σ if the maser disk is assumed to be circular). The largest potential source of such a difference in either error budget is the uncertainty in the distance modulus to the LMC; however, other identified potential sources of systematic error are great enough that a revision of the LMC distance is not required by our results.

A number of other tests of the extragalactic distance scale have been attempted in recent years using independent, physical or geometric methods; however, none of them currently have the same precision and direct applicability to recent Cepheid measurements that the maser distance to NGC4258 may provide in combination with our work. Observations of a ‘light echo’ from the supernova SN1987A have been used to place upper limits on the distance to the LMC; however, results differing by 10% in distance from each other have been obtained with this method^{23,24}, and it cannot place constraints on errors related to WFPC2 calibration or the effects of differences in chemical composition from the LMC. Interferometric radio observations of the expansion of SN1993J in M81 have been used to measure its distance with $\sim 15\%$ uncertainty under modest assumptions²⁵; the agreement with the HST Cepheid distance to its host galaxy is excellent²⁶. However, the uncertainty in the distance to this supernova exceeds the estimated systematic uncertainties in the Cepheid distance scale, and the Cepheid distance to M81 was obtained with a different instrument

(WF/PC) and photometric calibration techniques from those used for most Key Project galaxies. Finally, distances to several such galaxies have been obtained using the expanding photospheres method (EPM) on supernovae that they hosted, again with good agreement. However, all of those supernovae were abnormal and considered *a priori* to be poor candidates for the method used; therefore, there may be substantial systematic errors in those EPM measurements, and drawing strong conclusions from this agreement may be questionable²⁷.

Assuming that the maser distance and the estimates of its uncertainty are correct, we can derive the resulting absolute magnitudes of the Cepheids observed and obtain a potential recalibration of the Cepheid period–luminosity relation; the details of this analysis will be described elsewhere (J.A.N. *et al.*, manuscript in preparation). Such a calibration could be significantly less subject to systematic effects than one using ground-based observations of the LMC, as Cepheids in NGC4258 should have a chemical composition similar to those in galaxies used by the Key Project to calibrate other distance indicators and have been observed with the same instruments, filters, and parameter measurement techniques used in other HST studies. To a first approximation, extinction-corrected WFPC2 Cepheid results (obtained as in the Key Project) would be revised using the NGC4258 maser distance simply by subtracting

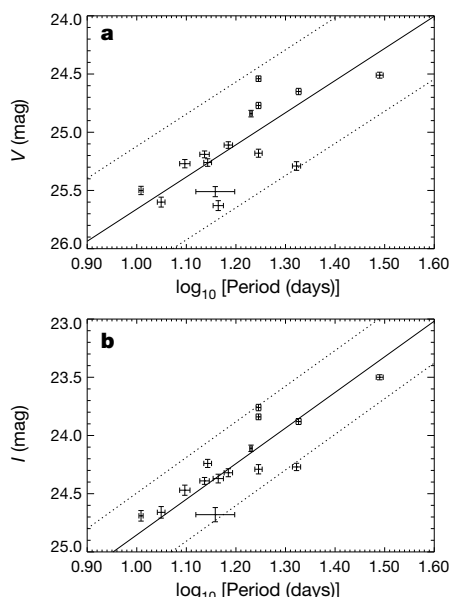


Figure 2 NGC4258 period–luminosity relations. **a**, The V period–luminosity ($P-L$) relation for Cepheids in NGC4258, based on DoPHOT photometry. The Cepheid $P-L$ relation has previously been found to be well fitted by a linear relationship between mean magnitude and the logarithm of the period in the range 10–60 days. To minimize the effects of incompleteness and to optimize results given the limited number of Cepheids observed, we adopted LMC $P-L$ relation slopes³ and only fitted for differences in the zero point. Shown are the measured parameters of NGC4258 Cepheids (symbols with error bars), the fitted $P-L$ relation (solid line), and the 2σ observed scatter of fiducial LMC Cepheids about its $P-L$ relation (dashed lines). From the difference between the absolute magnitudes of LMC Cepheids (for an assumed LMC distance of 50 kpc) and the observed magnitudes of NGC4258 Cepheids, we derive a V distance modulus [defined by $\mu = 5\log_{10}(d/10\text{pc})$] of $\mu_V = 29.82 \pm 0.07$ (ALLFRAME results yield 29.88 ± 0.08). This value must be corrected for the effects of dust extinction to yield a reliable measurement of the distance of NGC4258. **b**, The I $P-L$ relation for NGC4258 Cepheids. As for V , we derive an I distance modulus $\mu_I = 29.71 \pm 0.05$ (ALLFRAME photometry yields 29.78 ± 0.05). Correcting for the effects of dust on the apparent magnitude of each star using its observed colour, we obtain an extinction-corrected distance modulus of 29.54 ± 0.07 mag from DoPHOT photometry for all Cepheids; those on WFPC2 chip 2 alone yield 29.58 ± 0.12 , while those on chip 3 yield 29.51 ± 0.09 (for ALLFRAME photometry, the corresponding values are 29.64 ± 0.08 , 29.50 ± 0.12 , and 29.74 ± 0.10 mag).

0.25 ± 0.19 mag from their distance moduli. The chemical composition of the field studied in NGC4258 matches the average for the Key Project galaxies used to calibrate most secondary indicators; if distances to both are corrected for differences from the LMC and then revised so that the Cepheid distance to NGC4258 would match the maser distance, the net correction to the distance scale would thus remain 0.25 mag. The results for NGC4258 would thus imply that typical HST WFPC2 Cepheid distances may be too high by $12\% \pm 9\%$ (or $18\% \pm 9\%$ if the ALLFRAME results are adopted). If applied to all WFPC2 Cepheid distances, such a correction would eliminate the discrepancies between the Tully–Fisher relation calibrated using ground-based distances and that obtained using Key Project results²⁸, reducing its overall scatter accordingly. If such a comparison is fair, this suggests that the difference between the maser and Cepheid distances may be more likely to be due to the difficulty of WFPC2 calibration than to an error in the assumed LMC distance. However, selection effects or systematic errors in measuring the total galactic magnitudes of the ground-based calibrators, which are relatively large in angular extent, could also explain the Tully–Fisher discrepancy, because those total magnitudes are required for determining such distances. Revising Cepheid-calibrated distance indicators according to the maser distance would increase the measured Hubble constant by $12\% \pm 9\%$, and decrease the corresponding age of the Universe similarly. The resulting values would be increasingly difficult to reconcile with globular cluster ages unless the Universe has very low density or, particularly, if it has a non-negligible cosmological constant or similar negative-pressure component²⁹.

However, the statistical significance of such a revision remains limited. A more compelling test of the HST Cepheid distance scale based on the maser distance to NGC4258 would require a substantially larger sample of Cepheids (reducing the uncertainties in determining the $V-I$ period–luminosity relations and the reddening), and better determination of aperture corrections; these issues could be addressed simultaneously by searching for Cepheids with the HST in a field that contains more stars and has undergone more recent star formation. It would be reasonable to expect that observations of a region richer in Cepheids might yield three times as many Cepheids (giving a distance modulus uncertainty of 0.04 mag) and aperture corrections accurate to ± 0.04 mag; better agreement between ALLFRAME and DoPHOT analyses might also occur with improved data. Reductions of uncertainties in the maser distance (for example, by improving constraints on the eccentricity of the circumnuclear disk) would also greatly increase its usefulness in calibrating the extragalactic distance scale. Successful maser distances to other galaxies, establishing a Hubble relation, would more firmly establish this technique. With improvements in both Cepheid and maser analysis, NGC4258 could establish a new primary step in the distance ladder, reducing the potential systematic errors in measurements of the Hubble constant to perhaps as little as 5% . □

Received 13 July; accepted 12 August 1999.

1. Mould, J. R. *et al.* The HST Key Project on the Extragalactic Distance Scale XXVIII. Combining the constraints on the Hubble constant. *Astrophys. J.* (submitted).
2. Hernstein, J. R. *et al.* A geometric distance to the galaxy NGC4258 from orbital motions in a nuclear gas disk. *Nature* **400**, 539–541 (1999).
3. Madore, B. F. & Freedman, W. L. The Cepheid distance scale. *Publ. Astron. Soc. Pacif.* **103**, 933–957 (1991).
4. Jacoby, G. H. *et al.* A critical review of selected techniques for measuring extragalactic distances. *Publ. Astron. Soc. Pacif.* **104**, 599–662 (1992).
5. Freedman, W. L. in *Critical Dialogues in Cosmology* (ed. Turok, N.) 92–129 (World Scientific, Singapore, 1997).
6. Kennicutt, R. C. Jr, Freedman, W. L. & Mould, J. R. Measuring the Hubble constant with the Hubble Space Telescope. *Astron. J.* **100**, 1476–1491 (1995).
7. Saha, A. *et al.* Discovery of Cepheids in IC 4182: Absolute peak brightness of SN Ia 1937C and the value of H_0 . *Astrophys. J.* **425**, 14–34 (1994).
8. Freedman, W. L. in *Particle Physics and the Universe* (World Scientific, Singapore, in the press); also as preprint astro-ph/9905222 at (<http://xxx.lanl.gov>) (1999).
9. Freedman, W. L. in *Relativistic Astrophysics and Cosmology* (eds Olinto, A. V., Schramm, D. N. & Frieman, J. A.) (World Scientific, Singapore, 1998).

10. Miyoshi *et al.* Evidence for a black hole from high rotation velocities in a sub-parsec region of NGC4258. *Nature* **373**, 127–129 (1995).
11. Maoz, E. Dynamical constraints on alternatives to supermassive black holes in galactic nuclei. *Astrophys. J.* **494**, L181–L184 (1995).
12. Greenhill, L. J., Henkel, C., Becker, R., Wilson, T. L. & Wouterloot, J. G. A. Centripetal acceleration within the subparsec nuclear maser disk of NGC 4258. *Astron. Astrophys.* **304**, 21–33 (1995).
13. Stetson, P. B. The center of the core-cusp globular cluster M15: CFHT and HST observations, ALLFRAME reductions. *Publ. Astron. Soc. Pacif.* **106**, 250–280 (1994).
14. Schechter, P. L., Mateo, M. & Saha, A. DoPHOT, a CCD photometry program: Description and tests. *Publ. Astron. Soc. Pacif.* **105**, 1342–1353 (1993).
15. Ferrarese, L. *et al.* The Extragalactic Distance Scale Key Project. IV. The discovery of Cepheids and a new distance to M100 using the Hubble Space Telescope. *Astrophys. J.* **464**, 568–599 (1998).
16. Stetson, P. B. *et al.* The Extragalactic Distance Scale Key Project. XVI. Cepheid variables in an inner field of M101. *Astrophys. J.* **508**, 491–517 (1998).
17. Newman, J. A. *et al.* A Cepheid distance to NGC4603 in Centaurus. *Astrophys. J.* (in the press).
18. Cardelli, J. C., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
19. Ferrarese, L. F. *et al.* The HST Key Project on the Extragalactic Distance Scale XXVI. The calibration of Population II secondary distance indicators and the value of the Hubble constant. *Astrophys. J.* (in the press); also as preprint astro-ph/9908192 at (<http://xxx.lanl.gov>) (1999).
20. Madore, B. F. *et al.* The Hubble Space Telescope Key Project on the Extragalactic Distance Scale XV. A Cepheid distance to the Fornax cluster and its implications. *Astrophys. J.* **515**, 29–41 (1999).
21. Kennicutt, R. C. Jr *et al.* The Hubble Space Telescope Key Project on the Extragalactic Distance Scale. XIII. The metallicity dependence of the Cepheid distance scale. *Astrophys. J.* **498**, 181–194 (1998).
22. Zaritsky, D., Kennicutt, R. C. Jr, Huchra, J. P. H II regions and the abundance properties of spiral galaxies. *Astrophys. J.* **420**, 87–109 (1994).
23. Gould, A. & Uza, O. Upper limit to the distance to the Large Magellanic Cloud. *Astrophys. J.* **494**, 118–124 (1998).
24. Panagia, N. in *New Views of the Magellanic Clouds* (eds Chu, Y. H., Suntzeff, N. B., Hesser, J. E. & Bohlender, D.) (Astronomical Society of the Pacific, San Francisco, 1999).
25. Bartel, N. *et al.* Expansion of Supernova 1993J. *Bull. Am. Astron. Soc.* **194**, 6208 (1999).
26. Freedman, W. L. *et al.* The Hubble Space Telescope Extragalactic Distance Scale Key Project. I: the discovery of Cepheids and a new distance to M81. *Astrophys. J.* **427**, 628–655 (1994).
27. Eastman, R. G., Schmidt, B. P. & Kirschner, R. The atmospheres of type II supernovae and the expanding photosphere method. *Astrophys. J.* **466**, 911–937 (1996).
28. Sakai, S. *et al.* The Hubble Space Telescope Key Project on the Extragalactic Distance Scale XXIV. The calibration of Tully–Fisher relations and the value of the Hubble constant. *Astrophys. J.* (submitted).
29. Krauss, L. M. The end of the age problem, and the case for a cosmological constant revisited. *Astrophys. J.* **501**, 461–466 (1998).
30. Stetson, P. B. On the automatic determination of light-curve parameters for Cepheid variables. *Publ. Astron. Soc. Pacif.* **108**, 851–876 (1996).

Correspondence and requests for materials should be addressed to E.M.
(e-mail: maoz@ism.arc.nasa.gov).

Coupling strength of charge carriers to spin fluctuations in high-temperature superconductors

J. P. Carbotte*, E. Schachinger† & D. N. Basov‡

* Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada

† Institut für Theoretische Physik, Technische Universität Graz, A-8010 Graz, Austria

‡ Physics Department, University of California San Diego, La Jolla, California 92093-0319, USA

In conventional superconductors, the most direct evidence of the mechanism responsible for superconductivity comes from tunnelling experiments, which provide a clear picture of the underlying electron–phonon interactions^{1,2}. As the coherence length in conventional superconductors is large, the tunnelling process probes several atomic layers into the bulk of the material; the observed structure in the current–voltage characteristics at the phonon energies gives¹, through inversion of the Eliashberg equations, the electron–phonon spectral density $\alpha^2F(\omega)$. The situation is different for the high-temperature copper oxide superconductors, where the coherence length (particularly for *c*-axis tunnelling) can be very short. Because of this, methods such as optical spectroscopy and neutron scattering provide a better

route for investigating the underlying mechanism, as they probe bulk properties. Accurate reflection measurements at infrared wavelengths and precise polarized neutron-scattering data are now available for a variety of the copper oxides^{3–5}, and here we show that the conducting carriers (probed by infrared spectroscopy) are strongly coupled to a resonance structure in the spectrum of spin fluctuations (measured by neutron scattering). The coupling strength inferred from those results is sufficient to account for the high transition temperatures of the copper oxides, highlighting a prominent role for spin fluctuations in driving superconductivity in these materials.

There have been many suggestions for the mechanism involved in the superconductivity of the oxides. Although no consensus has yet emerged, the state of superconductivity is widely accepted to have *d*-wave symmetry^{6–9} with the gap on the Fermi surface vanishing along the main diagonals of the two-dimensional Brillouin zone. In YBCO there also exists extensive spin-polarized inelastic neutron-scattering data. These experiments reveal that spin excitations persist on a large energy scale¹⁰ (over several hundred meV), but are mainly confined around the (π, π) -point in momentum space. Also, in the superconducting state, a new peak emerges out of, or is additional to, the spin excitation background, which is often referred to as the 41 meV resonance⁵ peak (Fig. 1). This peak has received much attention but its origin remains uncertain^{11,12}. In one view, it is a result of a readjustment in the spin excitation spectrum due to the onset of superconductivity¹¹. Such a readjustment of spectral weight with a reduction below twice the superconducting gap value Δ_0 (refs 13, 14) is expected and is generic to electronic mechanisms. A second view is that it is a resonance in the $SO(5)$ unification¹² of magnetism and superconductivity.

If the spin excitations are strongly coupled to the charge carriers they should be seen in optical experiments. The normal state optical conductivity $\sigma(\omega)$ as a function of frequency ω depends on the electron self energy $\Sigma(\omega)$ which describes the effect of interactions on electron motion. In an electron–phonon system the electron–phonon interaction spectral density, $\alpha^2F(\omega)$, is approximately³ (but not exactly) equal to $W(\omega)$, a second derivative of the inverse of the normal state optical conductivity

$$\alpha^2F(\omega) \approx W(\omega) = \frac{1}{2\pi} \frac{d^2}{d\omega^2} \left[\omega \operatorname{Re} \frac{1}{\sigma(\omega)} \right] \quad (1)$$

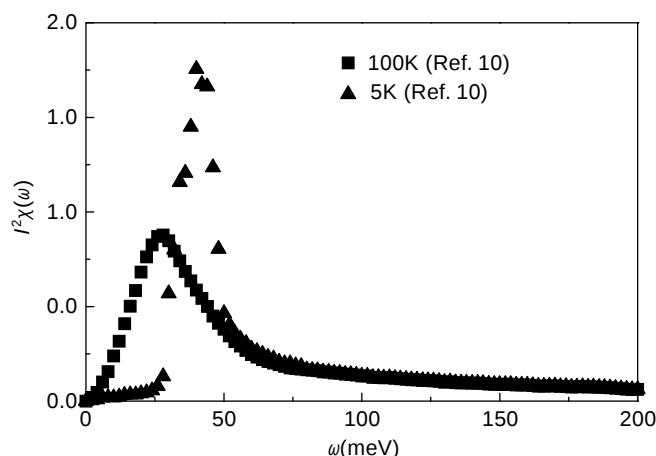


Figure 1 Spin-polarized inelastic neutron-scattering data for $\text{YBa}_2\text{Cu}_3\text{O}_{6.92}$. Neutron-scattering results taken at 100 K in the normal state (solid squares) compared with similar data taken at 5 K in the superconducting state (solid triangles) showing the 41 meV resonance. The data have been scaled by a constant coupling strength I^2 fixed to obtain $T_c = 100$ K.

In the phonon energy range, the correspondence is remarkably close and determines $\alpha^2 F(\omega)$ with good accuracy. At higher energies, additional, largely negative wiggles come into $W(\omega)$ which can simply be ignored as they are not part of $\alpha^2 F(\omega)$. Note that equation (1) is dimensionless and so determines the absolute scale of the electron-phonon interaction spectral density as well as its shape in frequency. This is important, as it allowed Marsiglio *et al.*³ to determine the $\alpha^2 F(\omega)$ of K_3C_{60} from its optical conductivity by inversion (1) and to conclude from a solution of the Eliashberg equations that it is large enough to explain the observed value of critical temperature. (T_c is related to the mass enhancement factor λ , twice the first inverse moment of $\alpha^2 F(\omega)$, ref. 2.)

The formalism for the normal state conductivity can also be applied to spin excitations. If we ignore anisotropy as a first approximation, we can proceed by introducing an electron-spin excitation spectral density denoted by $I^2\chi(\omega)$ with its scale set by the coupling strength to the charge carriers, I^2 , and $\chi(\omega)$ the imaginary part of the spin susceptibility measured in spin-polarized inelastic neutron-scattering experiments averaged over all momenta in the Brillouin zone. At low temperatures $\chi(\omega)$ contains the 41 meV resonance observed in the superconducting state. Here we use $\chi(\omega)$ directly from experimental results on a $YBa_2Cu_3O_{6.92}$ sample with $T_c = 91$ K and near optimum doping, for which results exist at the temperatures $T = 100$ K and $T = 5$ K (ref. 10), both properly calibrated in units of μ_B^2/eV (μ_B is the Born magneton) as shown

in Fig. 1. We multiply $\chi(\omega)$ at $T = 100$ K by a constant coupling I^2 fixed to get $T_c = 100$ K (ref. 15). The mass enhancement factor λ (twice the first inverse moment of $I^2\chi(\omega)$) obtained is 2.6 and is, to within 10%, the same as that obtained from the $W(\omega)$ derived from the normal state experimental data of Basov *et al.*¹⁶ in $YBa_2Cu_3O_{6.95}$, and from our calculated $W(\omega)$ at T_c . In a preliminary attempt to invert, Collins *et al.*¹⁷ found a λ of three which is greater than our value. Their twinned crystals exhibited a higher optical scattering rate than our untwinned crystal and consequently they obtained about 50% more weight in the main peak of $I^2\chi(\omega)$ around 30 meV.

In order to access lower temperatures, we need to understand how the $I^2\chi(\omega)$ structure enters the superconducting state optical conductivity. We have done a series of calculations of the superconducting state $\sigma(\omega)$ for a d -wave superconductor including inelastic scattering¹⁵. We used the method of ref. 15 to calculate the theoretical $\sigma(\omega)$, using the neutron data taken for $YBa_2Cu_3O_{6.92}$ (at 5 K) as $\chi(\omega)$, multiplied by the same value of the coupling strength I^2 that was previously determined to obtain a T_c of 100 K from the normal state neutron data. We then inverted this theoretical $\sigma(\omega)$ data using equation (1). The result of the inversion is compared in Fig. 2a (solid line) with our input spectral density $I^2\chi(\omega)$ (solid triangles) shifted in energy by the gap $\Delta_0 = 27$ meV of our theoretical calculations (manuscript in preparation).

The absolute scale of $I^2\chi(\omega)$ in the resonance region is well supplied by the peak value in the solid curve. This peak is followed by negative wiggles, which are not in the original input spectrum because $W(\omega + \Delta_0)$ is not exactly $I^2\chi(\omega)$. Nevertheless, this procedure allows us to see directly, by spectroscopic means, some of the features of $I^2\chi(\omega)$ and, more importantly, to obtain information on its absolute value at maximum. The long tails in $I^2\chi(\omega)$ at higher energy, extending well beyond the resonance, are not resolved in $W(\omega)$. Instead, they cause $\tau^{-1}(\omega)$, defined as $\text{Re}\{\sigma^{-1}(\omega)\}$, to rise in a quasi-linear fashion at high frequencies⁴ in both normal and superconducting state, as is observed. This quasi-linear rise was the motivation for the marginal Fermi liquid model¹⁴, which gives $\tau^{-1}(\omega) \propto \omega$ and a constant spectral density for $\omega > T$ extending to high energies. If we approximate the normal state experimental $\tau^{-1}(\omega)$ data⁴ at T_c by a straight line for $0 \leq \omega \leq 200$ meV, we obtain a value of 0.3 for the weight of the spectral density for all frequencies $\omega > T$ and a λ value of 2.8; these are quite consistent with our previous estimates. It is also important that, although the 41 meV resonance is near $2\Delta_0$, the density of quasiparticle states (not shown), has structure at about 50 meV in our calculations, a well established feature of tunnelling data.

In Figure 2b we show experimental results obtained¹⁶ on application of equation (1) to a -axis conductivity data on an untwinned single crystal. The 41 meV resonance is clearly resolved as a peak at approximately 69 meV in the solid curve (the gap is 27 meV). The height of this peak is ~ 3 , giving an absolute measure of the coupling between charge carriers and spin excitations. On comparison with Fig. 2a, we see that the coupling to the 41 meV resonance is larger experimentally than in the calculations that generated the theoretical results shown in Fig. 2a. This is not surprising as we have used the spin-polarized inelastic neutron-scattering data set measured on a near-optimum 91 K sample of $YBa_2Cu_3O_{6.92}$. The neutron results for slightly overdoped YBCO are very different⁵, although the T_c value is hardly affected. This large dependence of $\chi(\omega)$ on the sample can be used to argue against their role in establishing T_c . However, the function that controls the conductivity is a complicated weighting of the spin susceptibility involving details of the Fermi surface and points in the Brillouin zone away from (π, π) , as well as the coupling to the charge carriers. Thus, the correspondence between $I^2\chi(\omega)$ and $\chi(\omega)$ is complex. Optical experiments therefore reveal only that $I^2\chi(\omega)$ is not as strongly dependent on doping as is $\chi(\omega)$.

In Fig. 2b, we present experimental results for $W(\omega)$ in underdoped, untwinned $YBa_2Cu_3O_{6.6}$ (dashed line), compared with the

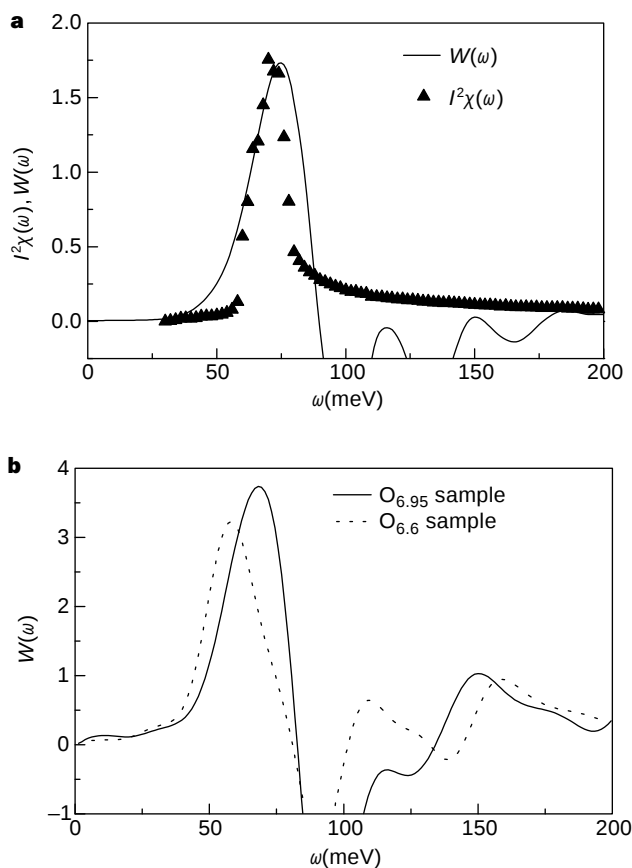


Figure 2 Inversion of the superconducting state optical conductivity. **a**, Results for $W(\omega)$ (solid curve) with $\sigma(\omega)$ calculated in the superconducting state at 5 K and based on the spectral density $I^2\chi(\omega)$ (solid triangles) based on the 5 K measured neutron data. In the region of the main resonance peak, the agreement with the input data, which has been shifted by the gap values (solid triangles), is excellent as to height and width. At higher energies, wiggles appear in $W(\omega)$ which are not present in $I^2\chi(\omega)$. **b**, $W(\omega)$ derived from the experimental data for the conductivity of an optimally doped $YBa_2Cu_3O_{6.95}$ single crystal with a -polarization (solid curve) and for an underdoped, untwinned $YBa_2Cu_3O_{6.6}$ single crystal (dashed line).

optimally doped case (solid line). It is interesting to note that the peak in the underdoped case is slightly reduced in height reflecting a reduction in T_c . It is also shifted to lower energies. Some experiments¹⁹ indicate a reduction in gap value with underdoping in YBCO while many experiments show an important increase in Bi2212 (ref. 18). Even if the gap is assumed to stay the same at 27 meV, the spin-polarized neutron-resonant frequency is known to decrease with doping²⁰. Accounting for this gives almost exactly the downward shift observed in our experimental data of Fig. 2b.

Recent inelastic neutron scattering data in Bi2212 (ref. 21) show a resonance peak at 43 meV in the superconducting state and establish a similarity with the earlier results in YBCO. We have inverted the optical data of Puchkov *et al.*⁴ in this case and find that coupling at low temperatures to the observed superconducting-state spin-resonance peak is a general phenomenon in both YBCO and Bi2212.

Spin excitations are thus seen in an appropriately chosen second derivative of the superconducting state optical conductivity, and hence the strength of their coupling to the charge carriers is determined. The coupling to the excitations including the 41 meV resonance is large enough in YBCO to account for superconductivity at that temperature. At T_c the spectrum obtained from experiment⁴ gives a value of the mass enhancement parameter λ which is close to the value used in our model calculations to obtain a critical temperature of 100 K.

Note added in proof: While this paper was in press, we became aware of a related theoretical study by Munzar, D. *et al.*, *Physica C* **312**, 121–135 (1999). □

Received 3 June; accepted 27 July 1999.

- McMillan, W. L. & Rowell, J. M. Lead phonon spectrum calculated from superconducting density of states. *Phys. Rev. Lett.* **14**, 108–112 (1965).
- Carbotte, J. P. Properties of boson exchange superconductors. *Rev. Mod. Phys.* **62**, 1027–1157 (1990).
- Marsiglio, E. *et al.* Inversion of K_2C_{60} reflectance data. *Phys. Lett. A* **245**, 172–176 (1998).
- Puchkov, A. V. *et al.* High T_c superconductors: an infrared study. *J. Phys. C*, **8**, 10049–10082 (1996).
- Bourges, P. in *The Gap Symmetry and Fluctuations in High Temperature Superconductors* (ed. Bok, J. *et al.*) 349–374 (Plenum, 1998).
- Scalapino, D. J. *et al.* d -Wave pairing near a spin-density-wave instability. *Phys. Rev. B* **34**, 8190–8192 (1986).
- Monthoux, P. & Pines, D. YBCO: a nearly antiferromagnetic Fermi liquid. *Phys. Rev. B* **47**, 6069–6081 (1993).
- Wollman, D. A. *et al.* Experimental determination of the superconducting pairing state in YBCO from the phase coherence of YBCO–Pb dc SQUIDS. *Phys. Rev. Lett.* **71**, 2134–2137 (1993).
- Tsuei, C. C. *et al.* Pairing symmetry and flux quantization in a tricrystal superconducting ring of YBCO. *Phys. Rev. Lett.* **73**, 593–596 (1994).
- Bourges, P. *et al.* Spin dynamics in high- T_c superconductors. (Preprint cond-mat/9902067).
- Bulut, N. & Scalapino, D. J. Neutron scattering from a collective spin fluctuation mode in a CuO_2 bilayer. *Phys. Rev. B* **53**, 5149–5152 (1996).
- Demler, E. & Zhang, S. C. Theory of the resonant neutron scattering of high- T_c superconductors. *Phys. Rev. Lett.* **75**, 4126–4129 (1995).
- Nuss, M. C. *et al.* Dynamic conductivity and ‘coherence peak’ in $YBa_2Cu_3O_7$ superconductors. *Phys. Rev. Lett.* **66**, 3305–3308 (1991).
- Varma, C. M. *et al.* Phenomenology of the normal state of $Cu-O$ high temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
- Schachinger, E. *et al.* Suppression of inelastic scattering on penetration depth and conductivity in $d_{x^2-y^2}$ superconductors. *Phys. Rev. B* **56**, 2738–2750 (1997).
- Basov, D. N. *et al.* Pseudogap and charge dynamics in CuO_2 planes in YBCO. *Phys. Rev. Lett.* **77**, 4090–4093 (1996).
- Collins, R. T. *et al.* Reflectivity and conductivity of $YBa_2Cu_3O_7$. *Phys. Rev. B* **39**, 6571–6574 (1989).
- DeWilde, Y. *et al.* Unusual strong coupling effects in the tunneling spectroscopy of optimally doped and overdoped $Bi_2Sb_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **80**, 153–156 (1998).
- Ponomarev, Yo. G. *et al.* Josephson effect and single particle tunneling in YBCO and YBCO single crystal break junctions. *Physica C* **243**, 167–176 (1995).
- Fong, H. F. *et al.* Superconducting-induced anomalies in the spin excitation spectra of underdoped YBCO. *Phys. Rev. Lett.* **78**, 713–716 (1997).
- Fong, H. F. *et al.* Neutron scattering from magnetic excitations in $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Nature* **398**, 588–591 (1999).

Acknowledgements

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institute for Advanced Research (CIAR). Work at University of California at San Diego (UCSD) is supported by the NFS ‘Early Career Development’ programme. DNB is a Cottrell Scholar of the Research Corporation. We thank J. E. Hirsch, P. B. Hirschfeld, P. B. Littlewood, F. Marsiglio, E. J. Nicol, D. Scalapino, T. Timusk, and I. Vekhter for interest.

Correspondence and requests for materials should be addressed to E.S. (e-mail: Schachinger@itp.tu-graz.ac.at).

Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98

Peter J. Webster*, Andrew M. Moore*, Johannes P. Loschnigg* & Robert R. Leben†

* Program in Atmospheric and Oceanic Sciences, Campus Box 311, University of Colorado, Boulder, Colorado 80309-311, USA

† Colorado Center for Astrodynamics Research, Department of Aerospace Engineering Sciences, University of Colorado, Boulder, Colorado 80309-311, USA

Climate variability in the Indian Ocean region seems to be, in some aspects, independent of forcing by external phenomena such as the El Niño/Southern Oscillation^{1–4}. But the extent to which, and how, internal coupled ocean–atmosphere dynamics determine the state of the Indian Ocean system have not been resolved. Here we present a detailed analysis of the strong seasonal anomalies in sea surface temperatures, sea surface heights, precipitation and winds that occurred in the Indian Ocean region in 1997–98, and compare the results with the record of Indian Ocean climate variability over the past 40 years. We conclude that the 1997–98 anomalies—in spite of the coincidence with the strong El Niño/Southern Oscillation event—may primarily be an expression of internal dynamics, rather than a direct response to external influences. We propose a mechanism of ocean–atmosphere interaction governing the 1997–98 event that may represent a characteristic internal mode of the Indian Ocean climate system. In the Pacific Ocean, the identification of such a mode has led to successful predictions of El Niño⁵; if the proposed Indian Ocean internal mode proves to be robust, there may be a similar potential for predictability of climate in the Indian Ocean region.

A strong, cool sea surface temperature (SST) anomaly (with respect to the 1950–97 mean as calculated from ref. 6) developed in the eastern Indian Ocean in July 1997 and reached a maximum ($>-2^\circ\text{C}$) in November 1997 (Fig. 1a). At about the same time, starting in June 1997, a warm SST anomaly developed in the western Indian Ocean, with a maximum of $>+2^\circ\text{C}$ in February 1998. Together, these heating and cooling anomalies produced a reversed SST gradient (SST increasing east to west) between November 1997 and June 1998 relative to the climatological temperature gradient (SST increasing west to east). Following the appearance of warm SST anomalies off the east African coast in June 1997, the usually weak climatological equatorial westerly winds (calculated from the NCEP/NCAR near-surface zonal wind component⁷) were replaced by surface easterly winds. The wind anomalies exceeded -5 m s^{-1} over the central equatorial Indian Ocean in December 1997 (Fig. 1b). Between July 1997 and May 1998, the sea surface height (SSH) was depressed substantially in the eastern basin, compared to the mean value as described in ref. 8, and generally higher in the west. Maximum height differences along the equator exceeded 30 cm between November 1997 and May 1998 (Fig. 1c). These variations in the climate of the Indian Ocean sector have been noted elsewhere⁹. Usually, Indian Ocean SST variations associated with El Niño are $\sim 0.5^\circ\text{C}$, but exhibit very different patterns to that noted in Fig. 1a (ref. 10).

The spatial distribution of deviations from the long-term mean SST, the outgoing longwave radiation (OLR, as a proxy for rainfall), the surface zonal wind velocity for November 1997, and the SSH field for the period November 20–30 1997, are shown in Fig. 2a–d (the long-term mean fields of SST, OLR and near-surface wind, see Figs 1–3 in Supplementary Information). There is a clear spatial structure in the anomalous SST fields, with extreme values in the

eastern and western regions of the basin. Compared to the long-term average, the near-surface anomalous zonal wind fields point strongly to the east ($>-6 \text{ m s}^{-1}$) with the core located just south of the Equator between 60°E and 100°E (Fig. 2c). Usually the zonal winds in autumn (September–November) are weak westerlies (see Fig. 3 in Supplementary Information). Fields of OLR measured by satellites provide the black-body radiating temperature of the atmosphere. Low OLR values indicate emissions from the upper troposphere and, thus, from deep convective clouds which are mainly responsible for rain in the tropics. High values of OLR, on the other hand, signal emission from the low troposphere and indicate an absence of deep convective clouds¹¹. Thus, the OLR distribution in Fig. 2b depicts greater than average rainfall in the

western and northwestern Indian Ocean and diminished rainfall in the east. Along the Equator there is an OLR gradient of $\sim 70 \text{ W m}^{-2}$ —roughly the same magnitude as between the western and eastern Pacific Ocean, but over half the horizontal distance. The OLR gradient is consistent with an atmospheric circulation having strong rising motion above the western Indian Ocean, strong surface flow to the west and subsidence in the east^{12,13}. Analyses of the NCEP/NCAR data substantiates the existence of the strong longitudinal cell¹⁴. The SSH field (Fig. 2d) shows lower values in the east, especially near Sumatra, and higher values in the west. To the south of the Equator and in the central and western Indian Ocean there is an elongated ‘ridge’ some 30 cm above adjacent ocean regions. The location and magnitude of the ridge is consistent

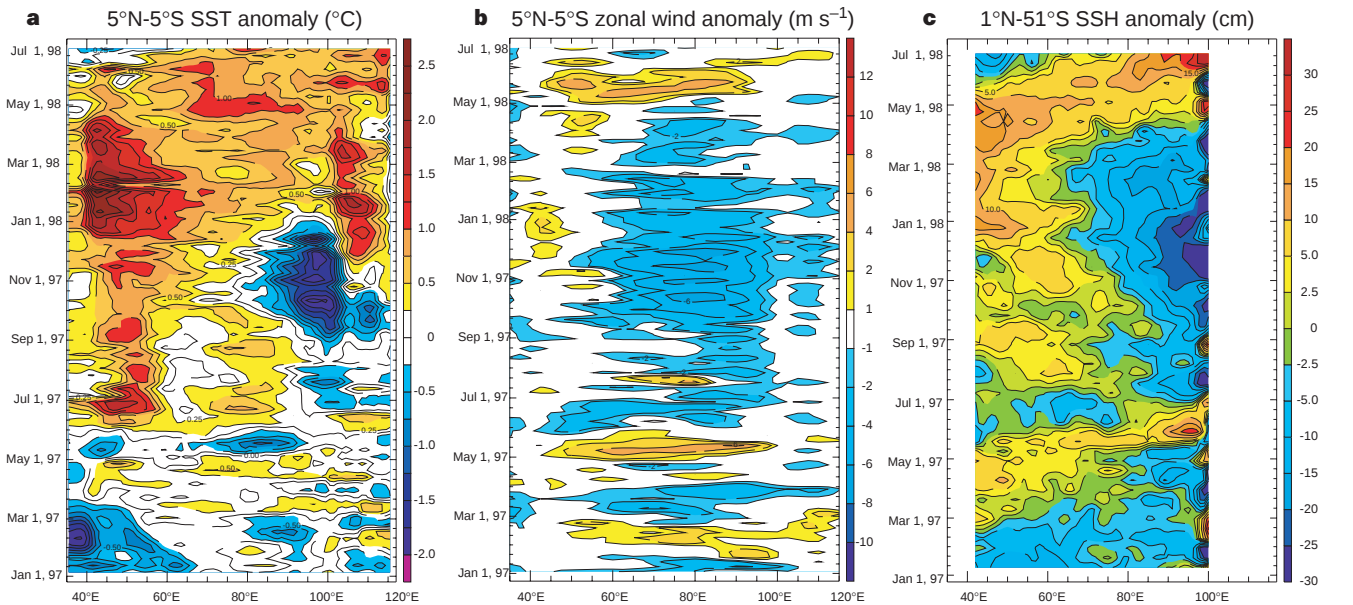


Figure 1 Evolution of the state of the Indian Ocean, 1997–98. Panels **a** and **b** show the anomalous SST and the zonal wind anomaly at the surface, respectively, averaged between 5°N and 5°S . The average annual cycle has been removed. Panel **c** shows the

SSH field obtained from TOPEX/POSEIDON averaged between 1°N and 1°S . The ordinate month marker denotes the beginning of the month.

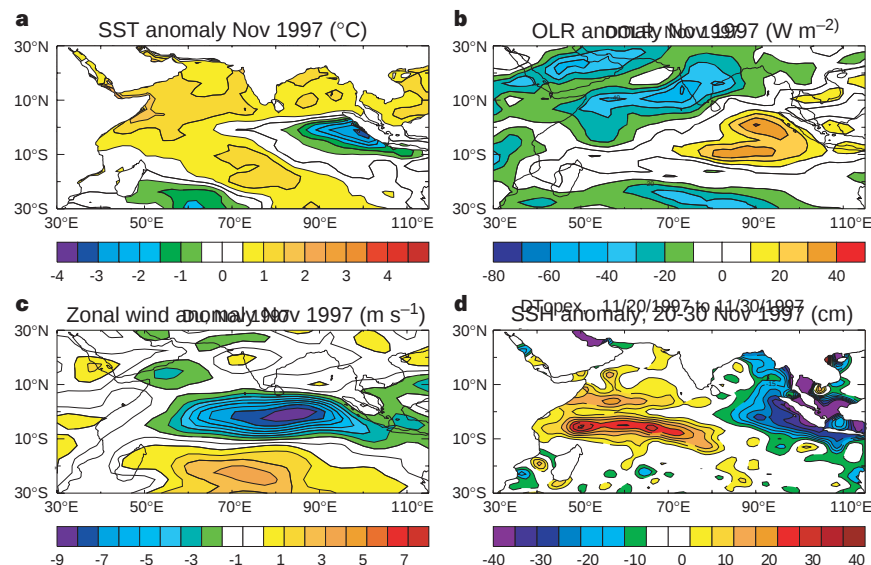


Figure 2 State of the Indian Ocean in November 1997. Panels show the latitude–longitude sections of anomalous (long-term monthly average removed) SST (**a**), outgoing longwave radiation (OLR) as determined from a satellite, which is a surrogate for

precipitation (**b**), the zonal component of the surface wind field (**c**), and the surface height determined from TOPEX/POSEIDON for 20–30 November 1997 (**d**).

with Southern Hemisphere Ekman mass transports to the left of the surface equatorial easterlies and westerlies near 10° S.

The time evolution of this ridge of elevated sea level south of the Equator is shown in Fig. 3. Longitude-time sections of SSH anomalies averaged between 4° and 6° S (Fig. 3a) and between 11° and 13° S (Fig. 3b) are shown for the period 1 January 1997 to 30 June 1998. Westward propagation of the ridge is evident at both latitudes, with lower westward phase speeds at higher latitudes. The westward phase speed and its decrease with increasing latitude suggests that ocean Rossby waves are important in the development and propagation of the ridge. Figure 3b also shows the emergence of Rossby waves from the eastern boundary around the time that winds are observed to increase along the coast of Sumatra, as discussed below. The timing, formation and the propagation of dynamical features of the Ekman ridge have been simulated with some accuracy by a simple one-and-a-half layer Gill-type model¹⁵ driven by observed surface wind fields. Analysis of the model results shows that the westward propagating modes are ocean downwelling Rossby waves.

Many of the changes observed in the SST and SSH fields can be explained by changes in the surface winds. Beginning in the late summer of 1997, the alongshore component of the surface winds off Sumatra were 2–3 m s⁻¹ stronger than normal to the northwest and were conducive to increased coastal upwelling, illustrated schematically in Fig. 4a. Starting in late summer, the alongshore winds off the African coast were weaker to the southwest by 2 m s⁻¹, which resulted in a reduction in upwelling south of the Equator. Initially, the warmer SSTs in the western Indian Ocean starting in June 1997 produced easterly wind anomalies over the central Indian Ocean that would later be enhanced by the cooler temperatures in the eastern basin. The accelerating equatorial easterlies would promote a large-scale dynamic adjustment of the SSH along the Equator, leading to a shallower eastern thermocline and a deeper western thermocline.

We suggest that the anomalous state in the Indian Ocean that we describe above may not have been an exaggerated response to

the strong 1997–98 El Niño. There are a number of reasons for this suggestion. First, the climate pattern around the Indian Ocean rim was very different during 1997–98 from that usually associated with a Pacific warm event. The drought in Indonesia was more intense than anticipated, the rainfall in the 1997 Indian summer monsoon and over Australia was near normal (although drought is usually expected in both regions), and the east African rainfall, usually only slightly higher during an El Niño, was the largest this century. Second, the typical El Niño/Southern Oscillation (ENSO) cycle appears more associated with a general warming of the equatorial Indian Ocean than a change in the longitudinal SST gradient, probably due to the lighter than average winds that occur over the Indian Ocean during an El Niño^{10,16}. The correlations, *R*, between the mean equatorial SST in the Indian Ocean and ENSO is +0.52, but falls to non-significant values of +0.19 between ENSO and the equatorial SST gradient. The correlation between October–November east African rainfall and the along-Equator Indian Ocean SST gradient is +0.62, indicating a relatively strong relationship between these two time series from within the Indian Ocean basin. Circulation and SST patterns in 1997–98 were similar to previous extremes in the east African rainfall (see Fig. 4a in Supplementary Information). The correlation between east African rainfall and the central Pacific SST anomalies, however, is only 0.24, indicating a rather weak relationship between east African rainfall and ENSO. In support of our findings, previous studies of African rainfall^{17,18} have suggested that other factors besides ENSO are responsible for east African rainfall variability. Third, similar excursions from normal have occurred in the Indian Ocean in the absence of ENSO extrema. During 1961–62, the SST gradient across the Indian Ocean reversed, with substantial warming in the western basin^{19,20}. In all, between 1950 and 1998 there were 16 years in which the equatorial SST gradient reversed for at least a month. Of these years, only three were El Niño years. None coincided with a La Niña. Spectral analysis of the time series of the along-Equator SST gradient (see Fig. 5 in Supplementary Information) shows peaks in a broad band around 2 years (exceeding 87% confidence limits) and 5 years (exceeding

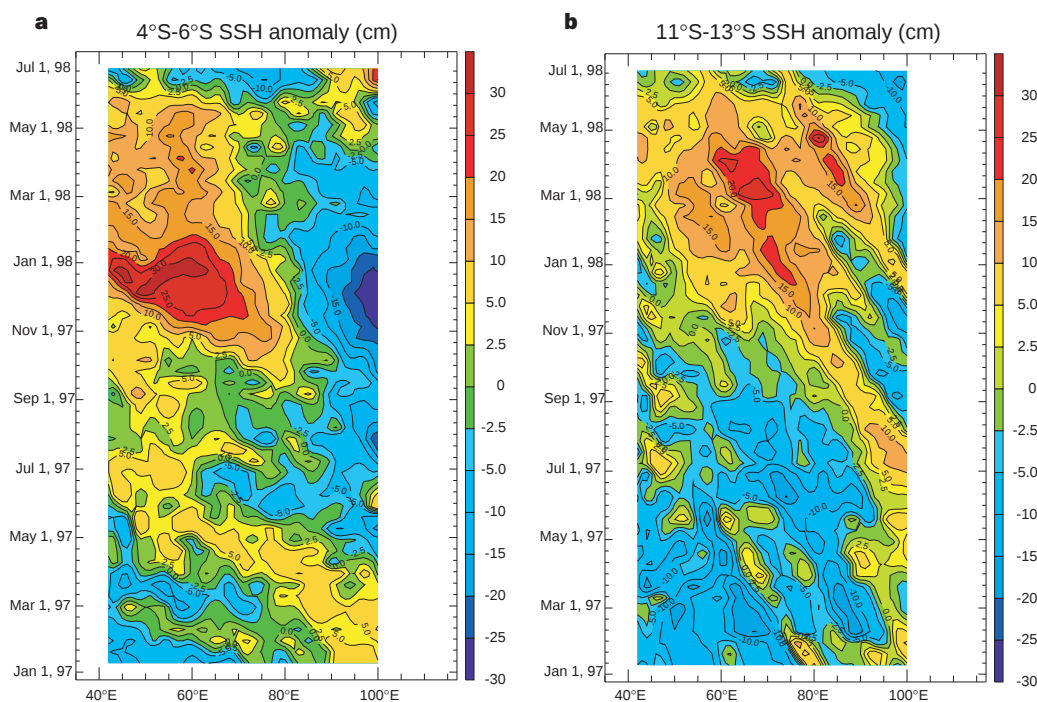


Figure 3 The evolving sea surface height (SSH) field. Panels show time–longitude sections of anomalous SSH obtained from TOPEX/POSEIDON averaged between 4° S and 6° S (a), and between 11° S and 13° S (b). Note the differential westward propagation as a

function of latitude. As latitude increases the phase speed decreases, consistent with the propagation characteristics of an ocean Rossby wave.

97%), indicating long-period oscillations in the gradient between two anomalous states: warm in the west and cool in the east, and cool in the west and warm in the east.

The possible sequence of events in the Indian Ocean occurring once the east–west equatorial SST gradient was established is shown in Fig. 4b–d. With a source of moist air from the Indian Ocean, strong convection develops over the heated land mass of east Africa during autumn of 1997, enhancing the equatorial easterly flow (Fig. 4b). Ekman transports, driven by the easterly anomalies, produce the Ekman ridge just south of the Equator. In seeking equilibrium, ocean Rossby wave dynamics cause the ridge to propagate westward which, in turn, deepens the thermocline to the west of the source of the Ekman convergence (Fig. 4c). With a deeper thermocline and reduced upwelling, the western Indian Ocean continued to warm, thus maintaining the driving force of

the anomalous easterly surface winds. In this manner, the Ekman ridge is sustained, allowing for a continual eastward propagation of the equatorial ocean Rossby waves. We propose that this self-sustaining system prolonged the warming in the western basin by a number of months.

A warming period may end, or diminish, in a number of ways. For example, a warm west Indian Ocean is often associated with a strong monsoon^{21,22}. This means that the summer monsoon following the Indian Ocean warming should have stronger winds in the western basin, which would induce greater mixing, greater Ekman transports forcing coastal upwelling, and greater evaporation, all of which would contribute to rapid cooling. Also, as the El Niño weakens (as it did during the spring of 1998), and the locus of convection moves back towards Indonesia, the equatorial wind patterns in the Indian Ocean revert to westerlies (Fig. 4d). The

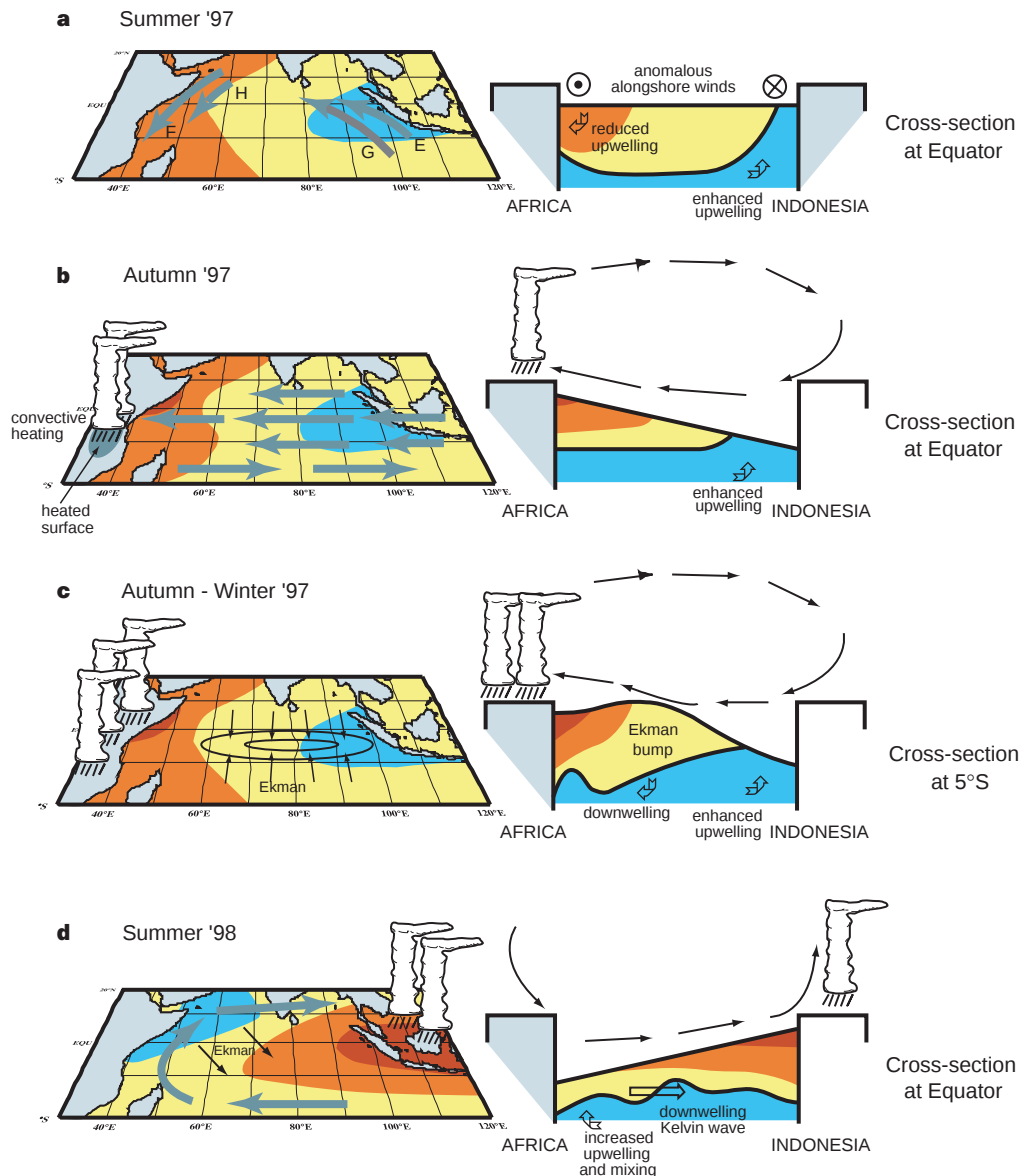


Figure 4 Diagram of the sequence of events in 1997–98. **a**, The climatological alongshore winds off Sumatra (E) and the east African coast (F). The winds observed in the late summer and early autumn are denoted by G and H, respectively. The right-hand panel shows the effect at the Equator on the upper ocean induced by increased upwelling in the east and decreased upwelling in the west. Wind into and out of the plane of the paper are denoted by the bull's eye and cross-hair symbols, respectively. **b**, Distribution of the winds resulting from the anomalous SST gradient along the Equator and the changes in the SSH

distribution. **c**, Formation of the Ekman ridge in the central Indian Ocean and the forcing of westward-propagating downwelling equatorial Rossby waves to the west. The right-hand panel shows the effect on the upper ocean near 5° S. **d**, Subsequent cooling of the western Indian Ocean through enhanced mixing and coastal Ekman transports from stronger than average monsoon winds and through circulation changes associated with the weakening of the 1997–98 El Niño.

resulting relaxation of the SSH fields would force eastward-propagating and downwelling Kelvin waves which would deepen the eastern mixed layer and return the system to a normal configuration. These changes may be seen in the spring and early summer of 1998 in Fig. 1.

The results presented above suggest that the Indian Ocean exhibits strong coupled ocean–atmosphere–land interactions that are self-maintaining, and are capable of producing significant perturbations to the annual cycle, at least during the 1997–1998 period. Arguably, the evolution of the perturbation is independent of ENSO. We note that other hypotheses^{3,4} have been suggested for observed interannual variability in the Indian Ocean. These theories rely on the Indian Ocean responding locally to either weaker or stronger monsoon winds which, through changes in upwelling and mixing, introduce a biennial component to the system. But the weakness of these theories is the maintenance of the upper-ocean anomalies from year to year. Our hypothesis adds a coupled dynamical component that has a longer timescale than the thermodynamics of the mixed layer, and which may form a link from one monsoon season to the next. Thus we suggest that the Indian Ocean may not be a passive player in climate variability on seasonal to interannual timescales, but may enact a very active and independent role. □

Received 6 July 1998; accepted 9 July 1999.

- Nicholls, N. Air-sea interaction and the quasi-biennial oscillation. *Mon. Weath. Rev.* **106**, 1505–1508 (1983).
- Nicholls, N. All-India summer monsoon rainfall and sea surface temperature around northern Australia and Indonesia. *J. Clim.* **8**, 1463–1467 (1995).
- Meehl, G. A. Coupled ocean-atmosphere-land processes and south Asian monsoon variability. *Science* **265**, 263–267 (1994).
- Meehl, G. A. The south Asian monsoon and the tropospheric biennial oscillation. *J. Clim.* **10**, 1921–1943 (1997).
- Webster, P. J. & Palmer, T. N. The past and future of El Niño. *Nature* **390**, 562–564 (1997).
- Reynolds, R. & Marisco, D. An improved real-time global sea surface temperature analysis. *J. Clim.* **6**, 114–119 (1993).
- Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Hendricks, J. R., Leben, R. R., Born, G. H. & Kobinsky, C. J. Empirical orthogonal function analysis of global TOPEX/POSEIDON altimeter data and implications for detection of global sea level rise. *J. Geophys. Res.* **101**, 14131–14145 (1996).
- Yu, L. & Rienecker, M. M. Mechanisms for the Indian Ocean warming during the 1997–1998 El Niño. *Geophys. Res. Lett.* **26**, 735–738 (1999).
- Barnett, T. P. Interaction of the monsoon and Pacific Ocean trade wind systems at interannual time scales. Part I: the equatorial zone. *Mon. Weath. Rev.* **111**, 756–773 (1983).
- Arkin, P. & Meisner, B. The relationship between large-scale convective rainfall and cloud cover over the western hemisphere during 1982–1984. *Mon. Weath. Rev.* **115**, 51–74 (1987).
- Webster, P. J. Response of the tropical atmosphere to local steady forcing. *Mon. Weath. Rev.* **100**, 518–541 (1972).
- Gill, A. E. Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **106**, 447–462 (1981).
- Near Real-Time Analysis of the Ocean and Atmosphere* Fig T 29, p. 36 (Climate Diagnostics Bull. NO. 97/11, Climate Diagnostics Center, National Center of Environmental Prediction, NOAA, Washington DC, 1997).
- Gill, A. *Atmosphere–Ocean Dynamics* (Academic, London, 1982).
- Webster, P. J. et al. Monsoons: processes, predictability and the prospects for prediction. *J. Geophys. Res.* **103**, 14451–14510 (1998).
- Nicholson, S. E. & Kim, J. The relationship of the El-Niño Southern Oscillation to African rainfall. *Int. J. Climatol.* **17**, 117–135 (1997).
- Nicholson, S. E. An analysis of the ENSO signal in the tropical Atlantic and western Indian Oceans. *Int. J. Climatol.* **17**, 345–375 (1997).
- Reverdin, G., Cadet, D. & Gutzler, D. Interannual displacements of convection and surface circulation over the equatorial Indian Ocean. *Q. J. R. Meteorol. Soc.* **122**, 43–67 (1986).
- Kapala, A., Born, K. & Flohn, H. In *Proc. Int. Conf. on Monsoon Variability and Prediction* (ed. Newson, R.) 119–126 (Tech. Doc. 619, World Meteorological Organization, Geneva, Switzerland, 1994).
- Rao, K. G. & Goswami, B. N. Interannual variations of the sea-surface temperature over the Arabian Sea and the Indian Monsoon: A new perspective. *Mon. Weath. Rev.* **116**, 558–568 (1988).
- Shukla, J. & Mooley, D. A. Empirical prediction of the summer monsoon over India. *Mon. Weath. Rev.* **115**, 695–703 (1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as hard copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the Office of Global Programs, NOAA, and the NSF (P.J.W., A.M.M., J.P.L.), and by NASA (R.R.L.).

Correspondence and requests for materials should be addressed to P.J.W. (e-mail: pjw@oz.colorado.edu).

A dipole mode in the tropical Indian Ocean

N. H. Saji*, B. N. Goswami†, P. N. Vinayachandran* & T. Yamagata*‡

* Institute for Global Change Research, SEAVANS N 7F, 1-2-1 Shibaura, Minato-ku, Tokyo 105 6791, Japan

† Center for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore 560 012, India

‡ Department of Earth and Planetary Physics, Graduate School of Science, The University of Tokyo, Tokyo 113 0033, Japan

For the tropical Pacific and Atlantic oceans, internal modes of variability that lead to climatic oscillations have been recognized^{1,2}, but in the Indian Ocean region a similar ocean–atmosphere interaction causing interannual climate variability has not yet been found³. Here we report an analysis of observational data over the past 40 years, showing a dipole mode in the Indian Ocean: a pattern of internal variability with anomalously low sea surface temperatures off Sumatra and high sea surface temperatures in the western Indian Ocean, with accompanying wind and precipitation anomalies. The spatio-temporal links between sea surface temperatures and winds reveal a strong coupling through the precipitation field and ocean dynamics. This air–sea interaction process is unique and inherent in the Indian Ocean, and is shown to be independent of the El Niño/Southern Oscillation. The discovery of this dipole mode that accounts for about 12% of the sea surface temperature variability in the Indian Ocean—and, in its active years, also causes severe rainfall in eastern Africa and droughts in Indonesia—brightens the prospects for a long-term forecast of rainfall anomalies in the affected countries.

The catastrophic rains of 1961 in tropical eastern Africa and subsequent abrupt discharge of the White Nile are now known^{4–6} to be part of an anomalous climate state over the tropical Indian Ocean. A dipole structure characterized the sea surface temperature (SST) anomaly during this event: warmer than usual SSTs occurred over large parts of the western basin, while SSTs off Sumatra were cooler than usual. Rainfall increased over tropical eastern Africa and the western Indian Ocean, while over the Indonesian archipelago it decreased, resulting in severe drought. Equatorial surface winds, which in a normal summer season blow towards the east, weakened and reversed direction. There was no El Niño in the Pacific, while India experienced the highest summer monsoon rainfall in the past 150 years (ref. 5). By examining long-term data sets of SST

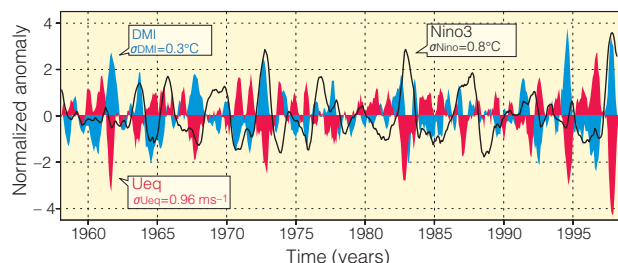


Figure 1 Dipole mode and El Niño events since 1958. Plotted in blue, the dipole mode index (DMI) exhibits a pattern of evolution distinctly different from that of the El Niño, which is represented by the Niño3 sea surface temperature (SST) anomalies (black line). On the other hand, equatorial zonal wind anomalies U_{eq} (plotted in red) coevolves with the DMI. All the three time series have been normalized by their respective standard deviations. We have removed variability with periods of 7 years or longer, based on harmonic analysis, from all the data sets used in this analysis. In addition, we have smoothed the time series using a 5-month running mean.

and surface winds we found several occurrences of these patterns. These observations motivated this study. However, here we use the GISST2.3b data set⁷ (1958–98), the NCEP surface winds reanalysis⁸ (1958–98) and the Xie–Arkin rainfall analyses⁹ (1979–98) to present our findings, as the essential signatures of this phenomenon are unchanged in the various data sets we have examined.

Basin-scale anomalies of uniform polarity cover³ the tropical Indian Ocean basin during El Niño/Southern Oscillation (ENSO) events. Using empirical orthogonal function (EOF) analysis we see that this pattern (EOF1) explains about 30% of the total variation of anomalous Indian Ocean SSTs. Next, the dipole mode (EOF2) explains about 12%. Characteristic to the dipole mode is a reversal in sign of SST anomaly across the basin. This reversal is so striking, that the dipole mode may be identified by a simple index time series which describes the difference in SST anomaly between the tropical western Indian Ocean (50° E–70° E, 10° S–10° N) and the tropical south-eastern Indian Ocean (90° E–110° E, 10° S–Equator). The strong correlation (>0.7) between this index, referred to as the dipole mode index (DMI), and the time series associated with EOF2 indicates the accuracy of the DMI in representing the dipole mode in SST.

The dipole mode event is independent of the ENSO in the Pacific Ocean. To demonstrate this, we plot SST anomalies representative of the central and eastern equatorial Pacific (from the so-called Nino3 region) against the DMI time series in Fig. 1. Note the significant dipole mode events of 1961, 1967 and 1994 coinciding with no ENSO, a La Niña and a weak El Niño, respectively. There are years in which dipole mode events coincide with strong ENSO events as in 1972 or 1997. We note here that the correlation between the DMI and Nino3 SST anomaly time series is weak (<0.35).

During dipole mode events, the surface wind field over the tropical Indian Ocean experiences large changes, especially in its zonal (east-to-west) component over the Equator. Maximum changes in the zonal wind occur over the equatorial central and eastern Indian Ocean, where we noted a correlation maximum of >|0.6| with the DMI. By plotting the area-averaged equatorial zonal wind anomalies (U_{eq}) over this region (70° E–90° E, 5° S–5° N), in Fig. 1, we demonstrate that the intensity of the SST dipole mode and the strength of the zonal wind anomaly over the Equator are strongly dependent on each other.

Seasonal phase locking is an important characteristic of the DMI time series. Thus significant anomalies appear around June, intensify in the following months and peak in October. Because of this systematic seasonality of phase it is meaningful to demonstrate the

evolution of a dipole mode event using composite analysis. We analysed the recent six extreme events of 1961, 1967, 1972, 1982, 1994 and 1997 to present the life cycle of a typical dipole mode event in Fig. 2. Cool SST anomalies first appear in the vicinity of the Lombok strait by May–June, accompanied by moderate south-easterly wind anomalies in the southeastern tropical Indian Ocean. In the following months, the cold anomalies intensify and appear to migrate towards the Equator along the Indonesian coastline, while the western tropical Indian Ocean begins to warm up. Zonal wind anomalies along the Equator and alongshore wind anomalies off Sumatra intensify together with the SST dipole. A dramatically rapid peaking of these features occurs in October, following by a rapid demise. The composite time series shown in Fig. 3 illustrates this sudden change of the DMI and U_{eq} more clearly. A strong feature in the time series is the tight coupling between the zonal wind anomaly and the dipole intensity. A biennial tendency—in this case a tendency for a dipole mode event to be preceded by an event of the opposite polarity—is evident. Note that the biennial tendency of U_{eq} is already well known¹⁰. Figure 3 also indicates that a pile-up of warm water off the Sumatran coast, together with westerly wind anomalies over the central-eastern equatorial Indian Ocean^{11,12}, signals the occurrence of a dipole mode event in the following year.

Basin-wide data for rainfall is available only since 1979. Analyses of the rainfall data (or proxy data sets such as outgoing longwave radiation) reveal that during a dipole mode event rainfall decreases over the oceanic tropical convergence zone (OTCZ) and increases over the western tropical Indian Ocean (Fig. 4). This pattern of rainfall is dynamically consistent with the convergence/divergence shifts associated with the wind field in the composite maps. Thus what emerges here is strong empirical evidence for coupling between the SST and the wind field through the precipitation field. Based on knowledge about the climate system in the Indian Ocean and from our own studies with a variety of modelling devices, we present the following model for how these fields are connected to each other.

In a normal year southeast trade winds converge into the South Equatorial Trough¹³ associated with the high-rainfall (>10 mm d⁻¹) OTCZ. When, during a dipole mode event, SSTs off Sumatra begin cooling, convection weakens at the OTCZ and the consequent surface pressure modification (not shown) makes the southeast trade winds extend and converge further downstream. This altered large-scale wind field enhances convergence and moisture supply at the extended downstream end of the trade winds, thereby

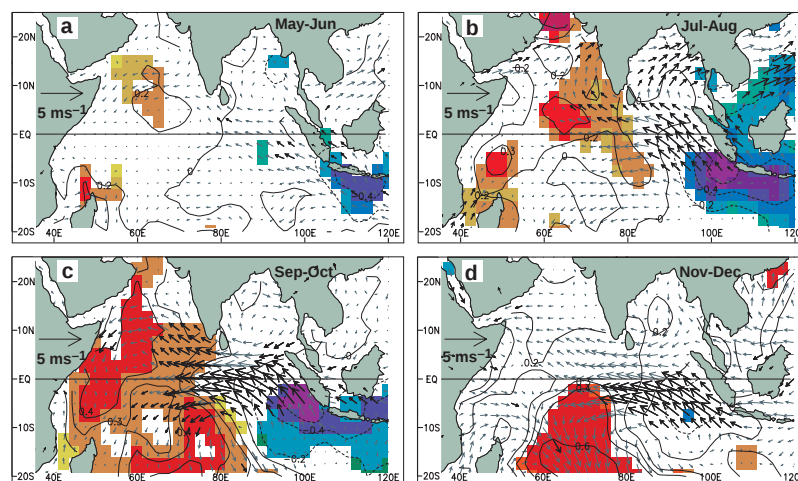


Figure 2 A composite dipole mode event. **a–d**, Evolution of composite SST and surface wind anomalies from May–June (**a**) to Nov–Dec (**d**). The statistical significance of the

analysed anomalies were estimated by the two-tailed *t*-test. Anomalies of SSTs and winds exceeding 90% significance are indicated by shading and bold arrows, respectively.

encouraging precipitation northwest of the normal position of the OTCZ. These abnormally extended trade winds also interrupt the normal heat supply to the coast off Sumatra. In a normal summer monsoon season, the westerly winds along the Equator accumulate warmer water along this coast through downwelling equatorial and coastally trapped Kelvin waves^{11,12,14}. The strongest manifestation of this process, occurring twice a year during monsoon transitions, is known as the Yoshida–Wyrtki Jet¹¹. This process counters the cooling tendencies by evaporation, coastal upwelling and oceanic heat advection brought about by the strong alongshore winds off this coast. In a year with a dipole mode event, abnormally extended trade winds with an easterly component along the Equator, by preventing the intrusion of the equatorial current^{15,16}, allow the cooling processes to dominate off Indonesia. This cooling is further enhanced as entrainment processes associated with the coastal winds become more effective due to the shallower thermocline^{17,18}. The shallowing thermocline is manifested in the observed lowering of the sea level^{16–20}. Hydrographic observations in this region also favour the above model of the coupling between surface winds and SST through ocean dynamics²¹. The warming in the west is initiated as a consequence of this chain of cause and effect taking place in the eastern half. The shifting trade winds, by increasing convergence in the west, lead to reduced wind speed and reduced evaporation^{18,20} which aid in warming up the SST. Entrainment is inhibited as increased rainfall raises the stability of surface waters through reduced salinity¹⁸. The thermocline anomalously deepens owing to reduced eastward transport^{11,22} resulting from the altered trade winds. The warmer SST increases the precipitation anomaly and consequently the wind anomaly to its east, thus introducing a positive feedback mechanism. The see-saw of the thermocline implied in the above discussion of the dipole mode mechanism is supported by the see-saw in both sea level^{16–20} and annual mean subsurface temperature anomalies²³ (not shown).

Here we have described a chain of events, which once initiated keeps the SST cooler in the tropical southeastern Indian Ocean, warmer in the tropical western Indian Ocean, and the southeast trade winds in the eastern half stronger than normal throughout the boreal summer and autumn. At this stage, the ocean–atmosphere system in the Indian Ocean approaches the configuration of those

in the Pacific and Atlantic oceans²⁴, and presumably would have attained a similar state had it not been for the rapid demise of the instability after the boreal autumn. Our preliminary understanding indicates that it is the changes in the state of the climate system, brought about by the seasonal monsoonal reversals, that are responsible for the demise of the dipole mode event. The main factor influencing the instability is the cooler-than-normal SST off Sumatra. Because the mass and heat transport from the west, achieved through equatorial ocean dynamics, play an important role in the heat balance of this region, its fluctuations affect the SST. However, after the boreal autumn, until the next spring, weakening winds both along the Equator and along the coast diminish the importance of ocean dynamics in the regulation of SST. Consequently, fluctuations of the mass and heat transport caused by equatorial dynamics play a lesser role in SST fluctuations. On the other hand, the increased insolation (as a result of the seasonal movement of the Sun) and reduced evaporation (due to decreasing windspeeds) dominate the changes in SST in this region during this season. In this context, it is likely that the higher-than-normal insolation, due to the reduction or absence of cloudiness in the tropical southeastern Indian Ocean, acting on the thin mixed layer, returns the system back to normality by removing the cool SST anomaly.

As the dipole mode is strongly dependent on the state of the system set up by the monsoonal circulation, it is expected that the variability of the monsoons would significantly affect this mode. We note that the dipole mode shares certain common features, such as the biennial tendency²⁵, with the monsoonal variability. Also, easterly anomalies along the Equator as well as reduced convection in the OTCZ are well-known features of strong monsoons²⁶. Nevertheless, the statistical correlation of the DMI with precipitation over the Asian monsoon regime does not yield a significant relationship (Fig. 4). Therefore the relationship of the DMI to the Indian monsoon variability is still not clear. It is clear, however, that the dipole mode has important implications for climate variability in other regions surrounding the Indian Ocean. The shift of the convergence zone, which leads to floods in east Africa and drought in Indonesia (Fig. 4), manifest the weakening or reversal of the zonal (Walker) circulation across the Indian Ocean. Besides causing this very discernible local climate variability across the basin, the zonal

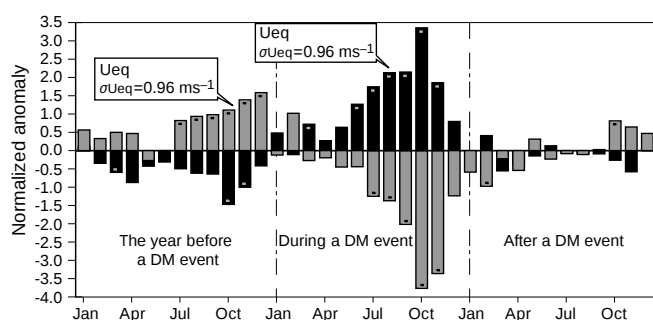


Figure 3 Strong coupling of SST dipole intensity to U_{eq} . Shown is the coevolution of intensity of the dipole mode (DMI, black bars) and strength of equatorial zonal winds anomalies (U_{eq} , grey bars) from the year before, to the year following a typical dipole mode event. Bars indicating significant anomalies (estimated by a two-tailed t -test) exceeding a 90% confidence level are marked with spots.

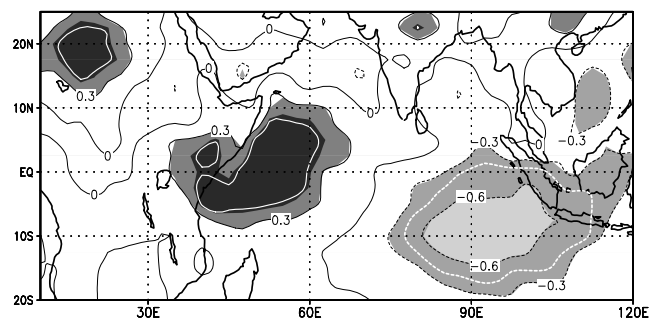


Figure 4 Rainfall shifts northwest of the OTCZ during dipole mode events. The map correlates the DMI and rainfall to illustrate these shifts. The areas within the white curve exceed the 90% level of confidence for non-zero correlation (using a two-tailed t -test).

redistribution of atmospheric mass and non-adiabatic heating during a dipole mode event may result in such an event influencing the countries surrounding the Indian Ocean through atmospheric teleconnections, as evidenced by the northward shift of the Pacific Subtropical High off the Philippines which caused an unusually hot summer in the far east Asian countries in 1994 (refs 27, 28). Some atmospheric general circulation model experiments^{29,30} have demonstrated the important effects of the Indian Ocean SST changes on east African rainfall variability. Further experiments with atmospheric and oceanic models, extended in scope, will be needed to understand the full implications of this hitherto unexplored mode of climate variability. The dipole mode event should also provide a testing ground for coupled ocean-atmosphere models. □

Received 27 April; accepted 27 July 1999.

- Neelin, J. D. *et al.* ENSO theory. *J. Geophys. Res.* **C 103**, 14261–14290 (1998).
- Zebiak, S. E. Air–sea interaction in the equatorial Atlantic region. *J. Clim.* **6**, 1567–1586 (1993).
- Wallace, J. M. *et al.* On the structure and evolution of ENSO-related climate variability in the tropical Pacific. *J. Geophys. Res.* **103**, 14241–14260 (1998).
- Flohn, H. East African rains of 1961/62 and the abrupt change of the White Nile discharge. *Palaeoecol. Afr.* **18**, 3–18 (1987).
- Kapala, A., Born, K. & Flohn, H. in *Proc. Int. Conf. on Monsoon Variability and Prediction* 119–126 (Int. Centre for Theoretical Physics, Trieste, 1994).
- Reverdin, G., Cadet, D. L. & Gutzler, D. Interannual displacements of convection and surface circulation over the equatorial Indian Ocean. *Q. J. R. Meteorol. Soc.* **112**, 43–67 (1986).
- Rayner, N. A., Horton, E. B., Parker, D. E., Folland, C. K. & Hackett, R. B. *Version 2.2 of the Global Sea Ice and Sea Surface Temperature Data Set, 1903–1994* (Clim. Res. Tech. Note 74, UK Meteorological Office, Bracknell, 1996).
- Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Xie, P. & Arkin, P. A. Analyses of global monthly precipitation using gauge observations, satellite estimates and numerical model predictions. *J. Clim.* **9**, 840–858 (1996).
- Barnett, T. P. Interaction of the Monsoon and Pacific Trade Wind System at interannual time scales. Part I: The equatorial zone. *Mon. Weath. Rev.* **111**, 756–773 (1983).
- Wyrtki, K. An equatorial jet in the Indian Ocean. *Science* **181**, 262–264 (1973).
- Yamagata, T., Mizuno, K. & Masumoto, Y. Seasonal variations in the equatorial Indian Ocean and their impact on the Lombok throughflow. *J. Geophys. Res.* **101**, 12465–12474 (1996).
- Stout, J. E. & Young, J. A. Low-level monsoon dynamics derived from satellite winds. *Mon. Weath. Rev.* **111**, 774–799 (1983).
- Clarke, A. J. & Liu, X. Observations and dynamics of semiannual and annual sea levels near the eastern equatorial Indian Ocean boundary. *J. Phys. Oceanogr.* **23**, 386–399 (1993).
- Reppin, J., Schott, F. A., Fischer, J. & Quadfasel, D. Equatorial currents and transports in the upper central Indian Ocean: Annual cycle and interannual variability. *J. Geophys. Res.* (in the press).
- Vinayachandran, P. N., Saji, N. H. & Yamagata, T. Response of the equatorial Indian Ocean to an unusual wind event during 1994. *Geophys. Res. Lett.* **26**, 1613–1616 (1999).
- Behera, S. K., Krishnan, S. & Yamagata, T. Anomalous air–sea coupling in the southern tropical Indian Ocean during the boreal summer of 1994. *Geophys. Res. Lett.* (in the press).
- Murtugudde, R., McCreary, J. P. & Busalacchi, A. J. Oceanic processes associated with anomalous events in the Indian Ocean with relevance to 1997–1998. *J. Geophys. Res.* (in the press).
- Potemra, J. & Lukas, R. Seasonal to interannual modes of sea level variability in the western Pacific and eastern Indian Oceans. *Geophys. Res. Lett.* **26**, 365–368 (1999).
- Yu, L. S. & Rienecker, M. M. Mechanisms for the Indian Ocean warming during the 1997–98 El Niño. *Geophys. Res. Lett.* **26**, 735–738 (1999).
- Meyers, G. Variation of Indonesian throughflow and the El Niño–Southern Oscillation. *J. Geophys. Res.* **101**, 12255–12263 (1996).
- Hastenrath, S., Nicklis, A. & Greischar, L. Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J. Geophys. Res.* **98**, 20219–20235 (1993).
- Levitus, S., Boyer, P. & Antonov, J. *World Ocean Atlas, Vol. 5, Interannual Variability of Upper Ocean Thermal Structure*. (NOAA Atlas NESDIS, NODC/NOAA, Silver Spring, Maryland, USA, 1994).
- Xie, S.-P. & Philander, S. G. H. A coupled ocean-atmosphere model of relevance to the ITCZ in the eastern Pacific. *Tellus A* **46**, 340–350 (1994).
- Goswami, B. N. A multiscale interaction model for the origin of the tropospheric QBO. *J. Clim.* **8**, 524–534 (1995).
- Goswami, B. N., Krishnamurthy, V. & Annamalai, H. A broad scale circulation index for the interannual variability of the Indian summer monsoon. *Q. J. R. Meteorol. Soc.* **125**, 611–633 (1999).
- Kawamura, R., Sugi, M., Kayahara, T. & Sato, N. Recent extraordinary cool and hot summers in east Asia simulated by an ensemble climate experiment. *J. Meteorol. Soc. Jpn.* **76**, 597–617 (1998).
- Nitta, T. Abnormal hot summer of 1994 and tropical convective activity. *Weath. Serv. Bull./Jpn. Meteorol. Agency* **63** (suppl.), 178–187 (1996) (in Japanese).
- Goddard, L. & Graham, N. E. The importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *J. Geophys. Res.* (in the press).
- Latif, M., Dommenget, D. & Dima, M. The role of Indian Ocean sea surface temperature in forcing east African climate anomalies. *J. Clim.* (in the press).

Acknowledgements

We thank H. Nakamura, Y. Masumoto, S. Behera and Y. Tanimoto for comments and discussions, and H. Watanabe and M. Yoshinaga for assistance in plotting the data.

Correspondence and requests for materials should be addressed to T.Y.

(e-mail: yamagata@geoph.s.u-tokyo.ac.jp).

Tribosphenic mammal from the North American Early Cretaceous

Richard L. Cifelli

Oklahoma Museum of Natural History and Department of Zoology,
University of Oklahoma, Norman, Oklahoma 73019, USA

The main groups of living mammals, marsupials and eutherians, are presumed to have diverged in the Early Cretaceous¹, but their early history and biogeography are poorly understood. Dental remains have suggested that the eutherians may have originated in Asia², spreading to North America in the Late Cretaceous, where an endemic radiation of marsupials was already well underway³. Here I describe a new tribosphenic mammal (a mammal with lower molar heels that are three-cusped and basined) from the Early Cretaceous of North America, based on an unusually complete specimen. The new taxon bears characteristics (molarized last premolar, reduction to three molars) otherwise known only for Eutheria among the tribosphenic mammals. Morphometric analysis and character comparisons show, however, that its molar structure is primitive (and thus phylogenetically uninformative), emphasizing the need for caution in interpretation of isolated teeth. The new mammal is approximately contemporaneous with the oldest known Eutheria from Asia. If it is a eutherian, as is indicated by the available evidence, then this group was far more widely distributed in the Early Cretaceous than previously appreciated. An early presence of Eutheria in North America offers a potential source for the continent's Late Cretaceous radiations, which have, in part, proven difficult to relate to contemporary taxa in Asia.

Class Mammalia

Infralegion Tribosphenida, Order and Family Uncertain

Montanalestes keebleri gen. et sp. nov.

Etymology. Montana: the state of origin; lestes (Greek): robber; *keebleri*: in honour of the Keebler family of Billings, Montana, in recognition of its support for field parties of the Oklahoma Museum of Natural History (OMNH).

Holotype. OMNH 60793, associated dentaries with right p3–5, m1–3 (terminology follows ref. 4), and alveoli for p2; and left p5, m1–3 (Fig. 1).

Locality and Horizon. OMNH V1156, basal unit VII (ref. 5), Cloverly Formation (Aptian–Albian; ~110 Myr); Carbon County, Montana, USA.

Diagnosis. Differs from primitive Tribosphenida (for example, *Kermackia*), in so far as is known, in having a semimolarized last premolar (lingually placed paraconid, incipient talonid basin, metaconid present) and only three molars; molars differ in having a broader, more fully basined talonid with three cusps consistently present. Differs from early marsupials and marsupial-like taxa⁶ in postcanine tooth formula and in having semimolarized last premolar and a smaller size differential between m1–2; lower molars differ in lacking a labial postcingulid and in having a greater trigonid–talonid height differential, narrower talonid and a hypoconulid that is not lingually situated. Differs from *Slaughteria*⁷ in having a semimolarized, less anteroposteriorly compressed last premolar; molars differ in having a smaller size differential between m1–2, no lingual knob on anteroconulid, more acute trigonid angle on m1 and taller paraconid relative to metaconid. Similar to early Eutheria in postcanine tooth formula (four or more premolars, three molars) and semimolarized last premolar; differs from most comparable taxon, *Prokennalestes*⁸, in having salient lingual cingulids on the lower premolars and stronger paraconid and metaconid on p5, the former more lingually placed. Lower molars differ from those of *Prokennalestes* in having sequential length

increase from first to third, taller paraconid, and greater height differential between trigonid and talonid and between protoconid and metaconid, more distinct and anteriorly placed entoconid, and hypoconulid more projecting on m1–2.

Description. The Meckelian groove (Fig. 2) and labial mandibular foramen, found in a number of Cretaceous mammals including presumed eutherians⁸, are absent. A roughened depression at the junction of the ascending and horizontal rami may indicate the presence of the coronoid, a presumed primitive condition with a broad distribution among early mammals, including some eutherians⁸. The angular process is elongate, hook-like, projects posteriorly and is in-turned medially, as in primitive Tribosphenida

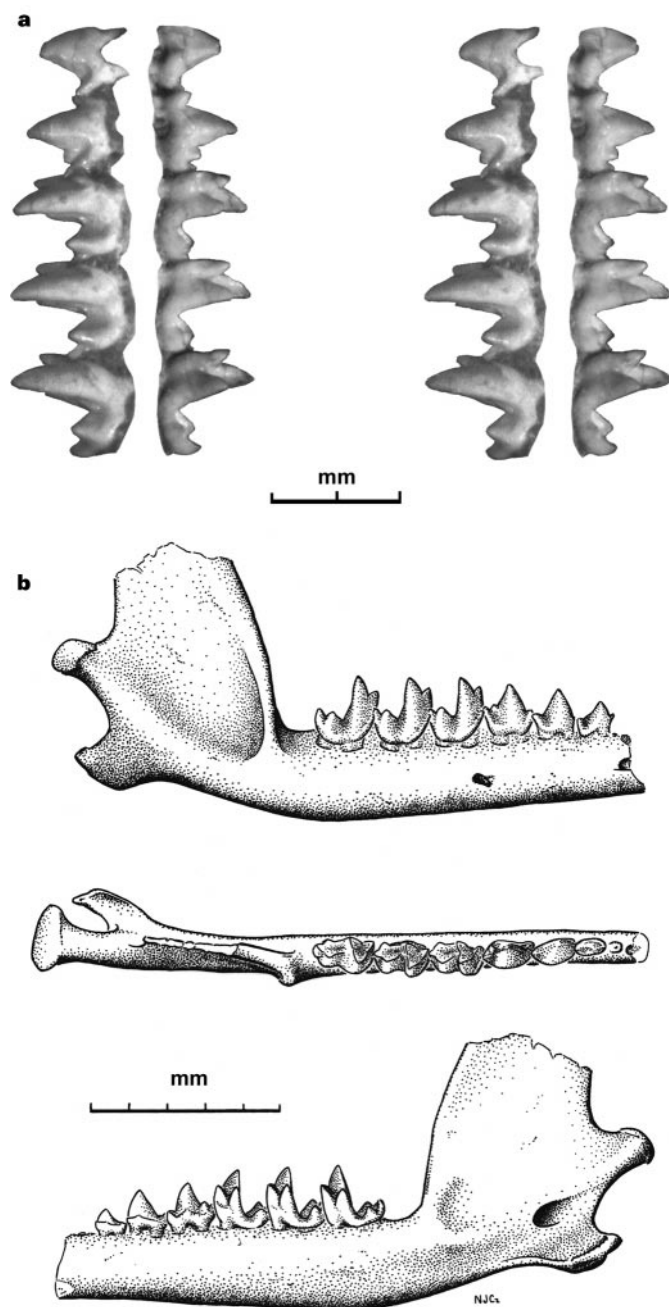


Figure 1 *Montanalestes keebleri* gen. et sp. nov., OMNH 60793. **a**, Stereopairs of p4–m3 (other parts of dentition and mandible omitted for clarity) in labial (left) and lingual (right) views. **b**, restored right mandible and dentition in labial (top), occlusal (middle) and lingual (bottom) views.

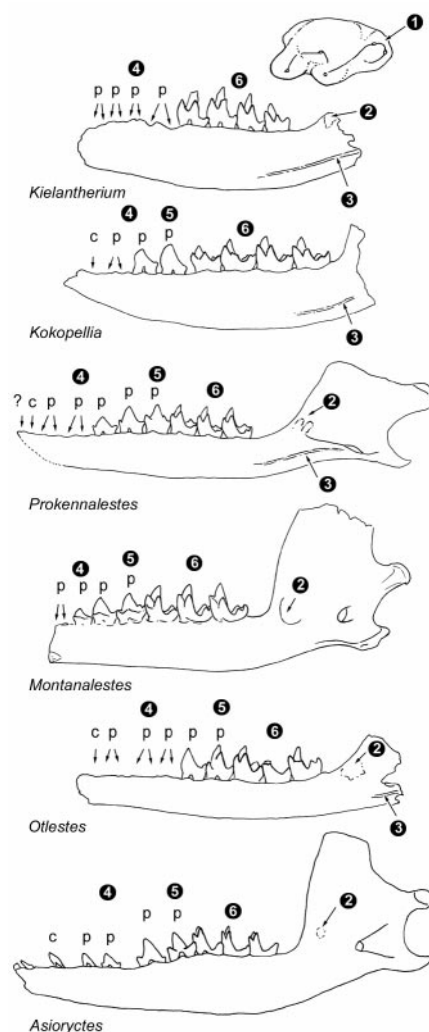


Figure 2 Jaws of Cretaceous tribosphenic mammals, showing distribution of selected tooth and jaw characters (some restored after refs given in Supplementary Information). *Kielantherium* is a plesiomorphic tribosphenidan of uncertain affinities¹⁴; *Kokopellia* is a marsupial or close relative⁶; and *Prokennalestes*⁸, *Otlestes*¹⁶ and *Asioryctes*²² are Cretaceous eutherians. 1, Talonid. Primitively (*Kielantherium*), this is small, poorly basined and bears only two cusps; in remaining taxa it is broader and well basined, with three cusps¹¹. 2, Coronoid bone. This element, or a facet indicating its presence, is a primitive feature presumably lost early in marsupial history and independently among eutherians, but retained in many Cretaceous mammals^{6,8}. 3, Meckelian groove (uncertain in *Otlestes*¹⁶ and *Kokopellia*⁶). This feature is another primitive character retained in some early eutherians and other Mesozoic mammals^{8,14,23}. 4, Premolar count. The primitive counts for premolars and molars are uncertain, depending partly on varying interpretation and relevance of non-tribosphenic *Peramus*, and partly on disputed homologies^{12,24}. At least four premolars are present in *Peramus* and *Kielantherium*, whereas deltatheroidans and marsupials have only three^{11,13–15}. The widespread occurrence of at least four in early eutherians is assumed to be primitive. 5, Last premolar. This tooth is simple in *Kokopellia*, marsupials and most Mesozoic mammals; by contrast, the premolars and molars of eutherians generally form a more evenly graded series, with the last premolar tending toward a semimolariform appearance^{8,12}. *Montanalestes* appears to be more advanced than the rather variable¹⁰ *Prokennalestes*, but the talonid and metaconid are not as well developed as in somewhat younger *Otlestes*¹⁶. 6, Molar count. This is uncertain for many early tribosphenidans⁷. *Kielantherium*, like marsupials and deltatheroidans, has four; *Montanalestes* is similar to eutherians in having only three molars, presumably an advanced condition^{11,12,15}. ?, alveolus for tooth of uncertain identity; p, premolar (or alveoli (arrows) documenting presence of a premolar).

generally⁹. At least four premolars are present; the relatively small size of p3 is similar to those in some Cretaceous eutherians that have five premolars¹⁰. The premolar–molar series intergrade: p5 is semimolariform, with a lingually placed paraconid; a metaconid that is clear but that is placed low on the crown and is not so lingually situated; and a small talonid basin enclosed lingually by the cingulum (although only one talonid cusp is present). The molars are virtually unworn and facets are poorly developed. They are advanced with respect to primitive tribosphenidans (Fig. 2) in having a well developed, basined, tricuspid talonid. The molars generally resemble those of pappotheriids and some early eutherians¹¹, although the latter tend to have a reduced paraconid¹². Molars of *Montanalestes* are unusual in having a relatively strong paraconid and a sequential increase in length from m1–3.

Montanalestes is undoubtedly a tribosphenic mammal, as shown by the broad, basined, three-cusped molar talonids¹³, and it is advanced with respect to dentally primitive taxa such as *Kielantherium*¹⁴ in various features (Fig. 2). *Montanalestes* is clearly not a marsupial: it retains more than three premolars, lacks specialized molar characteristics^{6,12,15} and has a molarized distal premolar^{11,12}, among other features. Eutherian affinities are, however, suggested by dental characteristics. The presence of four

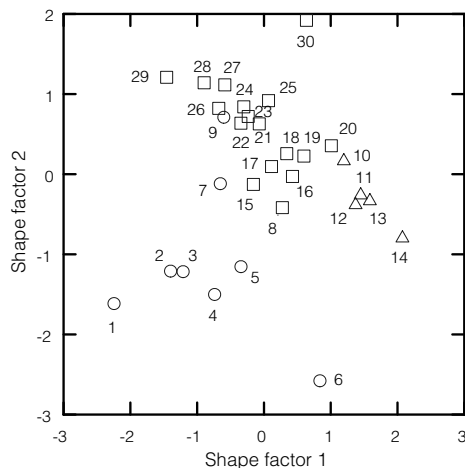


Figure 3 Comparison of lower molar shape in selected Cretaceous tribosphenic mammals and the eupantothere *Peramus*. Measurements and methodology follow ref. 11; data from this source has been standardized and variation due to size has been removed. Plot shows the first two axes of a principal components analysis. Positive scores (representing advanced conditions) on axis 1 reflect mainly increased length of the posthypocristid and width of the talonid, and relatively lower protoconid; this shape factor serves mainly to distinguish marsupials from eutherians and primitive tribosphenic/non-tribosphenic mammals. Positive scores on axis 2 reflect increased trigonid width, anteroposterior compression of the trigonid and lowered height differential between protoconid and metaconid. This shape factor generally distinguishes non-tribosphenic and some primitive tribosphenic mammals from the more advanced marsupials and, especially, eutherians. Taxa in the middle of the graph (for example, *Montanalestes* (8), *Pappotherium* (7), *Procerberus* (15)) have molars that, in terms of the two shape factors, are structurally antecedent to both marsupials and eutherians. Note placement and clustering of *Deltatheridium* (4) with primitive taxa (1–5), and placement of *Holoclemensia* (9) and *Kokopellia* (12) with respect to undoubted eutherians and marsupials. Circles, non-tribosphenic taxa and Theria of metatherian–eutherian grade; triangles, marsupials; squares, eutherians. 1, *Peramus*; 2, *Hypomylos*; 3, *Kielantherium*; 4, *Deltatheridium*; 5, *Kermackia*; 6, *Potamotelses*; 7, *Pappotherium*; 8, *Montanalestes*; 9, *Holoclemensia*; 10, *Didelphodon vorax*; 11, *Pediomys elegans*; 12, *Kokopellia*; 13, *Glasbius*; 14, *Alphadon marshi*; 15, *Procerberus*; 16, *Prokennalestes* spp.; 17, *Batodon*; 18, *Paranyctoides* spp.; 19, *Kennalestes*; 20, *Gallolestes pachymandibularis*; 21, *Gypsonictops hypoconus*; 22, *Cimolestes cereberoides*; 23, *Ottestes*; 24, *Asioryctes*; 25, *Purgatorius unio*; 26, *Cimolestes incisus*; 27, *C. magnus*; 28, *Protungulatum* spp.; 29, *Cimolestes propaleoryctes*; 30, *Zalambdalestes lechei*.

molars appears to be primitive for tribosphenic mammals^{11,12,15}; *Montanalestes* shares the condition seen in eutherians, a presumed reduction to three molars. The premolars of most early mammals are simple and form a sharp break with the molar series, whereas eutherians are presumed to be derived in having premolar–molar series intergrade, with a tendency for the last premolar to be somewhat molariform^{8,12}. *Montanalestes* is similar to eutherians in this respect: the last premolar bears a strong, lingually placed paraconid, a metaconid and a slightly basined talonid, all features commonly seen in premolars undergoing molarization. By comparison to *Prokennalestes* and other early Eutheria^{2,8,11,16}, *Montanalestes* presents a mixture of advanced (for example, metaconid on p5, loss of Meckelian groove) and primitive (for example, unreduced paraconid) features. Molar structure is of particular interest in this context because it has traditionally been emphasised in interpretation of early mammal radiations^{11,17}. Morphometric analysis shows that marsupials and dentally advanced, Late Cretaceous eutherians can be recognized in the basis of lower molar shape factors; plesiomorphic tribosphenidans are also distinct (Fig. 3). However, certain early tribosphenidans, *Montanalestes* and some early eutherians are structurally appropriate antecedents to both major groups of tribosphenic mammals. Lower molars appear not to be particularly informative for critical early tribosphenid taxa (as illustrated by the case of *Holoclemensia*, earlier suggested to be marsupial¹⁷, but more closely resembling eutherians based on referred lower molars; Fig. 3), indicating that taxa known only from molar structure must be interpreted cautiously. The relationship of *Montanalestes* to other early eutherians remains uncertain because available data produce ambiguous results.

The earliest generally accepted eutherians are from the Early and early Late Cretaceous of Asia^{8,10,16,18}. Previous reports of Early Cretaceous Eutheria in North America¹⁷ are based on extremely fragmentary specimens and have been controversial because of differing interpretations of tooth formula and specimen association⁷. Uncontested eutherians do not appear in North America until the early Campanian (about 80 Myr)¹⁹. By contrast, marsupials and presumed relatives were well established on the continent by the medial Cretaceous, about 100 Myr^{6,20}. These distributions have led to the suggestion that eutherians and marsupials arose in Eurasia and North America, respectively, and that North American eutherians are derivatives of Late Cretaceous immigration from Asia². Except for ungulates⁴, however, Late Cretaceous eutherians of Asia and North America represent distinct clades whose intercontinental relationships are problematic^{11,19}. If *Montanalestes* is related to Eutheria, as the balance of evidence indicates, then the continent of origin for the group is called into question: evidently, Eutheria enjoyed a far broader Early Cretaceous distribution (echoing that of the triconodont mammal *Gobiconodon*²¹) and an earlier evolutionary radiation than previously envisaged. An early presence of Eutheria in North America indicates that some of the continent's distinctive Late Cretaceous taxa may have evolved *in situ*, rather than having a proximal Asiatic origin—a theory to be tested by recovery of further faunas from the medial Cretaceous, where the record so far has been notoriously poor¹³. □

Received 12 April; accepted 27 July 1999.

1. Lillegraven, J. A., Thompson, S. D., McNab, B. K. & Patton, J. L. The origin of eutherian mammals. *Biol. J. Linn. Soc.* **32**, 281–336 (1987).
2. Butler, P. M. in *Structure, Function and Evolution of Teeth* (eds Smith, P. & Tchernov, E.) 125–138 (Freund, Tel Aviv, 1990).
3. Kielan-Jaworowska, Z. in *Palaeontology, Essential of Historical Geology* (ed. Gallitelli, E. M.) 367–383 (S.T.E.M. Mucci, Modena, 1982).
4. Nessel, L. A., Archibald, J. D. & Kielan-Jaworowska, Z. Ungulate-like mammals from the Late Cretaceous of Uzbekistan and a phylogenetic analysis of Ungulatomorpha. *Bull. Carnegie Mus. Nat. Hist.* **34**, 40–88 (1998).
5. Ostrom, J. H. Stratigraphy and paleontology of the Cloverly Formation (Lower Cretaceous) of the Bighorn Basin area, Montana and Wyoming. *Bull. Peabody Mus. Nat. Hist.* **35**, 1–234 (1970).
6. Cifelli, R. L. & de Muizon, C. Definition and jaw of *Kokopellia juddi*, a primitive marsupial or near marsupial from the medial Cretaceous of Utah. *J. Mamm. Evol.* **4**, 241–258 (1997).

7. Butler, P. M. A new interpretation of the mammalian teeth of triposphenic pattern from the Albian of Texas. *Breviora* **446**, 1–27 (1978).
8. Kielan-Jaworowska, Z. & Dashzeveg, D. Eutherian mammals from the Early Cretaceous of Mongolia. *Zool. Scripta* **18**, 347–355 (1989).
9. Novacek, M. J. The skull of lepidictid insectivores and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183**, 1–112 (1986).
10. Sigogneau-Russell, D., Dashzeveg, D. & Russell, D. E. Further data on *Prokennalestes* (Mammalia, Eutheria inc. sed.) from the Early Cretaceous of Mongolia. *Zool. Scripta* **21**, 205–209 (1992).
11. Butler, P. M. Early trends in the evolution of tribosphenic molars. *Biol. Rev.* **65**, 529–552 (1990).
12. Clemens, W. A. Jr & Lillegraven, J. A. New Late Cretaceous, North American advanced therian mammals that fit neither the marsupial nor eutherian molds. *Contrib. Geol. Univ. Wyoming* **3**, 55–85 (1986).
13. Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A. Jr (eds) *Mesozoic Mammals—the First Two-thirds of Mammalian History* (Univ. of California Press, Berkeley, 1979).
14. Dashzeveg, D. & Kielan-Jaworowska, Z. The lower jaw of an aegialodontid mammal from the Early Cretaceous of Mongolia. *Zool. J. Linn. Soc.* **82**, 217–227 (1984).
15. Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
16. Nessov, L. A., Sigogneau-Russell, D. & Russell, D. E. A survey of Cretaceous tribosphenic mammals from middle Asia (Uzbekistan, Kazakhstan and Tajikistan), of their geological setting, age and faunal environment. *Palaeover.* **23**, 51–92 (1994).
17. Slaughter, B. H. Mid-Cretaceous (Albian) therians of the Butler Farm local fauna, Texas. *Zool. J. Linn. Soc.* **50** (suppl.), 131–143 (1971).
18. Wang, Y. Q., Hu, Y.-M., Chow, M.-C. & Li, C.-K. in *Sixth Symposium on Mesozoic Terrestrial Ecosystems* (eds Sun, A.-L. & Wang, Y.-Q.) 221–228 (China Ocean, Beijing, 1995).
19. Fox, R. C. *Paranyctoides maleficus* (new species), an early eutherian mammal from the Cretaceous of Alberta. *Spec. Publ. Carnegie Mus. Nat. Hist.* **9**, 9–20 (1984).
20. Cifelli, R. L. et al. in *Vertebrate Fossils of Utah* (ed. Gillette, D. D.) 219–242 (Utah Geological Survey, Salt Lake City, 1999).
21. Jenkins, F. A. Jr & Schaff, C. R. The Early Cretaceous mammal *Gobiconodon* (Mammalia, Triconodontata) from the Cloverly Formation in Montana. *J. Vert. Paleontol.* **8**, 1–24 (1988).
22. Kielan-Jaworowska, Z. Evolution of the therian mammals in the Late Cretaceous of Asia. Part II. Postcranial skeleton in *Kennalestes* and *Asioryctes*. *Palaeontol. Pol.* **37**, 65–83 (1974).
23. Luo, Z. in *In the Shadow of the Dinosaurs—Early Mesozoic Tetrapods* (eds Fraser, N. C. & Sues, H.-D.) 98–128 (Cambridge Univ. Press, Cambridge, 1994).
24. Luckett, W. P. in *Mammal Phylogeny, volume 2—Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 182–204 (Springer, New York, 1993).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

I thank W. D. Maxwell and D. Smith for their collaboration in field activities, S. K. Madsen for fossil preparation, N. J. Czaplewski for illustration, M. Kyte for cooperation of the US BLM, W. Helmer and C. Nomee for the cooperation of the Crow Tribe, and P. Butler for various help. Funding was provided by the National Geographic Society and the NSF.

Correspondence and requests for materials should be addressed to R.L.C.

Oldest playable musical instruments found at Jiahu early Neolithic site in China

Juzhong Zhang^{*†}, Garman Harbottle[‡], Changsui Wang[†] & Zhaochen Kong[§]

^{*} Institute of Cultural Relics and Archaeology of Henan Province, Zhengzhou, Henan, China 450000

[†] Archaeometry Laboratory, University of Science and Technology of China, Hefei, Anhui, China 230026

[‡] Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA

[§] Paleobotany Laboratory, Institute of Botany, Academia Sinica, Beijing, China 100080

Excavations at the early Neolithic site of Jiahu^{1,2} in Henan Province, China have produced what may be the earliest complete, playable, tightly-dated multinode musical instruments. Jiahu was occupied from 7000 BC to 5700 BC, considerably antedating the well known Peiligang culture^{3–5}. Here we describe six exquisitely made complete flutes which were found in radiocarbon-dated excavation layers, along with fragments of perhaps 30 more. The

flutes are made from the ulnae of the red-crowned crane (*Grus japonensis* Millen) and have 5, 6, 7 and 8 holes. The best preserved flute has been played and tonally analysed. In addition to early musical artefacts, the archaeological record at Jiahu^{1,2} contains important information on the very foundations of Chinese society. We describe the archaeological characteristics of the Jiahu site, details concerning its dating, its place in the prehistory of the Chinese Neolithic, the ethnicity of its population and the results of a tonal analysis of a nearly 9,000-year-old musical instrument found there.

Jiahu lies in the 'Central Yellow River Valley'³ in mid-Henan Province, east of Mount Funiu and bounded by the flood plains of the Ni river to the south and the Sha river to the north. It was discovered by Zhu Zhi in 1962 but its significance was not fully appreciated until other sites related to the Peiligang culture^{3–5} had been excavated. Jiahu is an irregular ellipse, 275 m long (east–west) and 260 m wide (north–south). Below a modern agricultural layer of loess 0.3–1.0 m deep lies a layer of the Han Dynasty (206 BC–AD 220), and beneath this Neolithic layers which together are 0.5–0.9 m thick. Of Jiahu's 55,000 m², only 2,400 m² (4.4%) have been excavated, revealing 45 house foundations, 370 cellars, 9 pottery kilns, more than 300 graves and thousands of artefacts made out of bone, pottery, stone and other materials. There are three sub-periods at Jiahu; during the latest two the site shows distinct areas for residences, workshops and graves¹. The differential furnishing of the graves provides evidence for early social stratification. Jiahu appears to represent an organized, well established sedentary village of long duration in the very early Chinese Neolithic.

The dating of Jiahu is crucial. So far we have determined twenty ¹⁴C dates: nine from charcoal (three in each of the three sub-periods mentioned above), five from plant ash, four from human bone, one from a whole fruit pit and one from a broken grain of carbonized rice. The conventional ¹⁴C ages range from 7,017 to 8,285 ¹⁴C years BP. After dendrochronological correction the corresponding range is 7,700 to 9,000 calendar years BP or nearly 5700 to 7000 cal. (calibrated) BC. The three sub-periods are dated as follows: the oldest is 6600–7000 cal. BC; the second oldest 6200–6600 cal. BC; and the third ~5700–6200 cal. BC. Thus, Jiahu spans some 1,300 years. If this range of dates is borne out by subsequent research, then Jiahu represents a Neolithic cultural phase distinctly earlier at its inception than the Peiligang culture, where the dates of many of the sites range from 5100 BC to 6300 BC^{3–5}. On the basis of these dates, and considering the repertoire of ceramic shapes and stone tools¹, we would tentatively characterize Jiahu as a culture parallel to, overlapping in duration, and very possibly related to, the Peiligang. The same might apply to the site of Houli in the eastern Yellow River Valley (north-central Shandong Province), dated 5000 BC to 6200 BC³. For the ¹⁴C determinations, see Supplementary Information.

A careful study of the bones of 400 individuals removed from more than 300 graves indicates that the Jiahu ethnic group may be identified with the North Asian Mongolian group, and also with the Xiwanggang and Miaodigou groups in Henan Province and the Dawenkou, Yedian and Xixiahou groups that were later found in Shandong Province. The range of male heights was from 170 to 180 cm. In the late Palaeolithic Zhoukoudian Cave, 'unspecialized'

Table 1 Analysis of intervals of flute M282:20

Location of interval	Average value in cents ¹⁴	Musical description
Between hole 1 and hole 2	284	Minor third
Between hole 2 and hole 3	244	Slightly larger than major second (a whole tone)
Between hole 7 and tube sound	260	Slightly smaller than minor third, slightly larger than a whole note.



Figure 1 Bone flutes from burials at Jiahu. Top to bottom: M341:2, M341:1, M78:1, M253:4, M282:20, M282:21. Scale is in centimetres.

Mongoloids were described⁶. By the Yangshao period (3000 BC–5000 BC)⁷, the skull measurements are ‘physically Chinese’ and ‘modern’⁶. The physical similarity of the Jiahu people to the later Dawenkou (2600 BC–4300 BC) indicates that the Dawenkou might have descended from the Jiahu, following a slow migration along the middle and lower reaches of the Huai river and the Hanshui valley.

The discovery of complete, playable multinote flutes at Jiahu presents us with a rare opportunity to hear and analyse actual musical sounds as they were produced nine millennia ago. Earlier flutes have been found in Neanderthal contexts, but they are so fragmentary that it is difficult to do more than guess their tonal production^{8–10}. We arranged to test one or more of the six flutes discovered at Jiahu; all are of the vertically held type (Sachs–Hornbostel¹¹ classification 421.111.12).

The best preserved flute (M282:20; Fig. 1), which was free of cracks, was chosen to be tested using a ‘Stroboscopp’ sound-analysing stroboscope, supervised by Huang Xiangpeng from the Music School of the Art Institute of China. This flute has seven main holes plus a tiny hole near hole 7. Two other seven-holed flutes were considered, but playing tests produced cracking sounds and were promptly discontinued. However, data were recorded for two players blowing twice each with their embouchures angled at 45° up and 45° down across the mouth of flute M282:20 (eight scales altogether).

The music research team did not use the modern standard of A4 = 440 Hz, but instead adopted an arbitrary standard of hole 5 = ‘C6’. (Based on A4 = 440 Hz, the actual tone of hole 5 was C6 + 2 Hz (± 20 Hz), averaged over eight trials.) Then the interval relationships of the sounds from hole 3 to hole 7 fitted reasonably well to the note sequence E6, D6, C6, B5, A5, with the tone of hole 1 = A6 and hole 2 = F#6. On this scale, the tone of the whole tube is G5 or F#5. In Table 1 three of the intervals in M282:20 are evaluated numerically.

Tests revealed that the tiny hole next to hole 7 (Fig. 1) was probably drilled to correct the off-pitch tone of the original hole 7; thus a tone of G#5 + 16 Hz was corrected to A5 – 11 Hz, which is much closer to the octave of A6 – 36 Hz.

Without testing more flutes, we cannot say whether the tonal scale of the bone flute of Jiahu (M282:20) is the ancestor of either the six-tone Qing Shan scale or the seven-tone Xia Shi scale; in any

case, the latter two scales are only documented six millennia later. It should be possible, by constructing exact replicas of the Jiahu flutes in material whose density approximates bird-bone, to study the tonal sequences of all these instruments without endangering the valuable artefacts themselves. The carefully selected tone scale observed in M282:20 indicates that the Neolithic musician of the seventh millennium BC could play not just single notes, but perhaps even music. It is important in considering the possible role of these flutes in Neolithic society to recall that ancient Chinese tradition held that there were strong cosmological connections with music: that music is part of nature¹². In this context, the performance of rituals and music were specifically associated with matters of state and sound government¹³.

Excavation of only a small fraction (<5%) of the Jiahu site has revealed that, by the unexpectedly early date of 7000 BC, a complex, highly organized Chinese Neolithic society had already begun to evolve employing multinote musical instruments. Future excavation and research should help us to understand the technical aspects of one of mankind’s earliest practices of musical expression, which probably took place in a ritual setting.

Note added in proof: Flute M282:20 can be heard on the *Nature* web site. In the recording, made at the Music Institute of the Art Research Institute of China, Taoying Xu plays part of the folk song “Xiao Bai Cai” (“The Chinese small cabbage”). Recording engineer was Bobao Gu. Research for the recording was by Xiangpeng Huang (deceased), Xinghua Xiao and Zhongliang Tong. □

Received 15 February; accepted 30 July 1999.

- Henan Province, Institute of Cultural Relics. Preliminary report for the second through the sixth excavation at the Neolithic site of Jiahu in Wuyang, Henan. *Wenwu* 1, 1–17 (1989).
- Zhang, J. & Wang, X. Notes on the recent discovery of ancient cultivated rice at Jiahu, Henan Province: a new theory concerning the origin of *Oryza japonica* in China. *Antiquity* 72, 897–901 (1998).
- Underhill, A. Current issues in Chinese Neolithic archaeology. *J. World Prehist.* 11, 103–160 (1997).
- An, Z. Radiocarbon dating and the prehistoric archaeology of China. *World Archaeol.* 23, 193–200 (1991).
- National Bureau of Cultural Relics Board. *A Compilation of Cultural Areas in China* (Zhongguo Ditu Press, Henan, 1991).
- Howells, W. W. in *The Origins of Chinese Civilization* (ed. Keightley, D. N.) (Univ. of California Press, Berkeley, 1983).
- Barnard, N. in *The Origins of Chinese Civilization* (ed. Keightley, D. N.) (Univ. of California Press, Berkeley, 1983).
- Marcuse, S. *A Survey of Musical Instruments* (Harper & Row, New York, 1975).
- Early Music. *Science* 276, 205 (1997).
- Lau, B., Blackwell, B. A. B., Schwarcz, H. P., Turk, I. & Blickstein, J. I. B. Dating a flautist? Using ESR

(electron spin resonance) in the Mousterian cave deposits at Divje Babe I, Slovenia. *Geoarchaeology* 12, 507–536 (1997).

11. Brown, H. M. in *The New Grove's Dictionary of Music and Musicians* (ed. Sadie, S.) 664–681 (Macmillan, London, 1987).
12. Pian, R. C., Kishibe, S. & Yang, B. N. in *The New Grove's Dictionary of Music and Musicians* (ed. Sadie, S.) 245–283 (Macmillan, London, 1987).
13. Needham, N. J. T. M., Wang, L. & Robinson, K. G. in *Science and Civilization in China* iv/1 (ed. Needham, N. J. T. M.) 126–228 (Cambridge University Press, 1962).
14. Lindley, M. in *The New Grove's Dictionary of Music and Musicians* (ed. Sadie, S.) 277–279 (Macmillan, London, 1987).

Supplementary Information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This project was supported by the National Natural Science Foundation of China. In addition, C.W. was supported by the Department of Science & Technology of China, the Chinese Academy of Sciences and the Structure Research Laboratory at USTC. Research at Brookhaven National Laboratory is supported by the US Department of Energy. We thank Huang Xiangpeng of the Music School of the Art Institute of China who supervised these important tests and the personnel of the same Music School who carried them out: Xiao Xinghua, Xu Taoying, Gu Bobao, Tong Zhongliang, Qiu Ping and Liu Haiwang.

Correspondence should be addressed to G.H. (e-mail: garman@bnl.gov) and requests for materials should be addressed to J.Z. For further information see http://www.pubaf.bnl.gov/pr/news_releases.html

Incorporating rules for responding into evolutionary games

John M. McNamara*, Catherine E. Gasson† & Alasdair I Houston†

* School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK

† School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

Evolutionary game theory^{1,2} is concerned with the evolutionarily stable outcomes of the process of natural selection. The theory is especially relevant when the fitness of an organism depends on the behaviour of other members of its population. Here we focus on the interaction between two organisms that have a conflict of interest. The standard approach to such two-player games is to assume that each player chooses a single action and that the evolutionarily stable action of each player is the best given the action of its opponent. We argue that, instead, most two-player games should be modelled as involving a series of interactions in which opponents negotiate the final outcome. Thus we should be concerned with evolutionarily stable negotiation rules rather than evolutionarily stable actions. The evolutionarily stable negotiation rule of each player is the best rule given the rule of its opponent. As we show, the action chosen as a result of the negotiation is not the best action given the action of the opponent. This conclusion necessitates a fundamental change in the way that evolutionary games are modelled.

Most two-player game models with a continuous range of possible actions assume that: (1) each player makes its choice before it has observed the action of its opponent; (2) a player cannot change its action once the opponent's action has been observed; (3) all population members that have a given role in the game are identical. For example, if each member of a pair of animals chooses the proportion of time it spends scanning for predators as opposed to feeding, then each would prefer to feed as much as possible and so would prefer the other to be vigilant. In models of this conflict^{3–5}, all the above assumptions hold; in particular, assumption (3) holds because the models assume (sometimes implicitly) that all animals are identical in terms of both energy reserves and ability to detect and escape from predators. If popula-

tion members follow an evolutionarily stable strategy, assumption (3) implies that every player in a given role adopts the same action, so that each player effectively knows the action of its opponent before this action is observed. Thus, given assumption (3), assumption (2) is reasonable. It is, however, not reasonable to assume that players in a given role are identical. Individuals will differ in aspects of quality that influence the costs and benefits of taking an action, and hence the action chosen will depend on quality. If the quality of an opponent cannot be directly observed, then an opponent's action is not known in advance and it will be beneficial to respond to this action once it has been observed. The opponent should respond in turn, and so on, until both players reach a negotiated settlement.

It might be thought that, if the variation in quality is small, this variation, and hence negotiation, can be ignored. We show, however, that this approximation may give quantitatively different predictions even when the variation in quality is vanishingly small.

We present an analysis of negotiation in the context of a pair of animals feeding their young, but the qualitative conclusions apply to a wide class of games. If a parent increases its feeding effort this will reduce its own future reproductive success, but will increase the success of itself and its partner in the current breeding attempt. Thus a conflict of interest exists, with each parent preferring the other to work hard. There is evidence that a parent responds directly to the effort of its partner^{6,7}, so that a model involving negotiation seems appropriate. In contrast, the standard model of parental effort by Houston and Davies⁸ involves a single decision. This game can be summarized as follows. All members of a given sex are identical. Each parent makes a single choice of effort, ignoring the partner's effort, so that there is no negotiation. If the male provides effort u_m and the female provides effort u_f then $B(u_m + u_f)$ young survive to maturity. These efforts reduce the future reproductive success of the male and the female by $K_m(u_m)$ and $K_f(u_f)$, respectively. For a given female effort u_f , the best male effort $\hat{r}_m(u_f)$, is the value of u_m that maximizes

$$B(u_m + u_f) - K_m(u_m) \quad (1)$$

The best female effort given that of the male, $\hat{r}_f(u_m)$, is defined

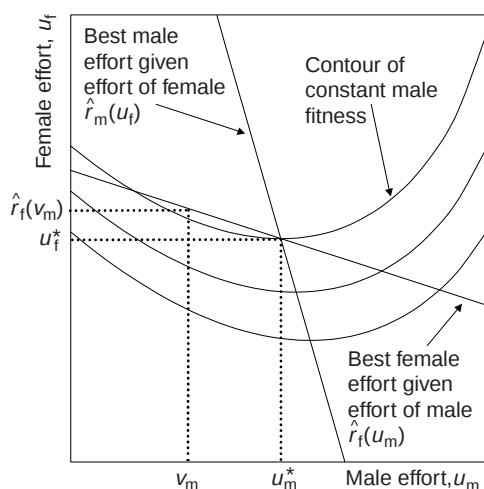


Figure 1 The best effort of a male, $\hat{r}_m(u_f)$, given fixed female effort u_f , and the best effort for the female, $\hat{r}_f(u_m)$, given fixed male effort u_m . The evolutionarily stable levels of effort, u_m^* and u_f^* , are the coordinates of the intersection of these two functions. The contours are lines of constant fitness for the male. For a given male effort, fitness increases with female effort. Assume that the female responds to the male with response rule $\hat{r}_f$. Then if a male provides effort v_m , the female responds with $\hat{r}_f(v_m)$ and the point $(v_m, \hat{r}_f(v_m))$ is above the contour that passes through (u_m^*, u_f^*) . Thus the male's fitness is higher than if he had used fixed effort u_m^* or had used $\hat{r}_m$ as a response rule.

analogously. The evolutionarily stable male effort, u_m^* , and female effort, u_f^* , satisfy

$$u_m^* = \hat{r}_m(u_f^*) \quad \text{and} \quad u_f^* = \hat{r}_f(u_m^*) \quad (2)$$

Thus, the effort of every male is best given the effort of its partner and vice versa. The functions $\hat{r}_m$ and $\hat{r}_f$ and the resulting evolutionarily stable levels of effort are shown in Fig. 1.

The functions $\hat{r}_m$ and $\hat{r}_f$ specify the response over evolutionary time of one sex to the fixed effort of the other sex. The functions have been interpreted as specifying how one pair member should behaviourally respond to the observed effort of its partner^{6–11}. This interpretation is incorrect, as then the population would not be evolutionarily stable. To see this, suppose that all population members adopted these behavioural response rules. The male and female members of a pair would then alternate their responses until their efforts u_m^* and u_f^* satisfied equation (2). The total reproductive success of a single mutant male that uses fixed effort u_m would be

$$B(u_m + \hat{r}_f(u_m)) - K_m(u_m) \quad (3)$$

This male can exploit the response rule of the female by taking his effort to be just less than u_m^* , as the female will then partially compensate for his lack of effort by increasing her own effort above u_f^* . Figure 1 shows that, as a consequence, the mutant male's reproductive success is greater than if he had used the response rule $\hat{r}_m$ (or had used fixed effort u_m^*).

In contrast to the assumptions of the model of ref. 8, parents will differ in quality (for example, foraging ability), which will influence the rate at which they should feed their young. Thus, as parents have been observed to respond to each other^{6,7}, a realistic model must seek evolutionarily stable negotiation rules rather than evolutionarily stable levels of effort. It is not straightforward to find such rules in general, so we concentrate on the following special model.

A negotiation involves the members of the breeding pair alternating in their choice of effort in response to that of their partner until the efforts of both partners settle down to limiting values. We refer to these final efforts as the outcome of the negotiation. We assume that in this process the effort adopted by an individual of quality q in response to the effort u of its partner is a function $r(u, q)$ of u and q alone, and does not depend on efforts earlier on in the negotiation. We refer to the function $r(u, q)$ as the response rule of

the individual. If a male of quality q_m has response rule $r_m(u, q_m)$ and a female of quality q_f has rule $r_f(u, q_f)$, then the outcome of negotiation will be efforts $\hat{u}_m$ and $\hat{u}_f$, respectively, where $\hat{u}_m = r_m(\hat{u}_f, q_m)$ and $\hat{u}_f = r_f(\hat{u}_m, q_f)$.

The negotiation phase is assumed to be cost free. If the outcome of negotiation is that a parent of quality q expends effort u , its loss in future reproductive success is $K(u, q)$, irrespective of its sex. An individual of high quality pays less of a cost for a given level of effort than an individual of low quality; specifically, $\frac{\partial^2 K}{\partial u \partial q} < 0$. By convention we set quality $q = 0$ to be that implicit in the model of ref. 8. Then q in our current model can be interpreted as a quality deviation. The evolutionarily stable efforts predicted by Houston & Davies⁸ are then given by $u_m^* = u_f^* = u^*$, where $B'(2u^*) = \frac{\partial K}{\partial u}(u^*, 0)$. If the quality deviations of both individuals are small, then it is reasonable that the final efforts are close to u^* . We can then Taylor-expand B to give the approximation

$$B(u_{\text{tot}}) = B(2u^*) + (u_{\text{tot}} - 2u^*)B'(2u^*) + \frac{1}{2}(u_{\text{tot}} - 2u^*)^2 B''(2u^*) \quad (4)$$

where $u_{\text{tot}} = u_m + u_f$. Our model takes this formula to be exact. By similar reasoning we take

$$K(u, q) = K + (u - u^*) \frac{\partial K}{\partial u} + q \frac{\partial K}{\partial q} + \frac{1}{2} \left[(u - u^*)^2 \frac{\partial^2 K}{\partial u^2} + 2q(u - u^*) \frac{\partial^2 K}{\partial q \partial u} + q^2 \frac{\partial^2 K}{\partial q^2} \right] \quad (5)$$

where K and its derivatives are evaluated at $u = u^*$ and $q = 0$.

As we show, the above form of B and K allows us to restrict attention to response rules of the form

$$r(u, q) = u^* + \rho + \mu q - \lambda(u - u^*) \quad (6)$$

where ρ , μ and λ are constants with $|\lambda| < 1$. Here ρ indicates bias from the baseline effort u^* and μ determines the dependence of an individual's effort on its own quality. The crucial parameter is the responsiveness λ , which measures the degree to which an individual's effort compensates for the lack of effort of its partner. The condition $|\lambda| < 1$ ensures that the successive efforts of a parent in the negotiation phase settle down to limiting values.

Suppose that almost all members of a population use a response rule of the form of equation (6) with the same constants ρ , μ and λ . Consider a single mutant individual that uses a negotiation rule that maximizes its reproductive success with every possible partner that it could have. We do not restrict the mutant to have a rule of the form of equation (6), and allow the mutant to base its efforts in the negotiation process on the whole history of the process before the present time. We show in Box 1 that the dependence of the final effort of this individual, $\tilde{r}(u, q)$, on its own quality and the final negotiated effort, u , of its partner is

$$\tilde{r}(u, q) = u^* + \tilde{\rho} + \tilde{\mu} q - \tilde{\lambda}(u - u^*) \quad (7)$$

where $\tilde{\rho} = f(\lambda)$, $\tilde{\mu} = g(\lambda)$ and $\tilde{\lambda} = h(\lambda)$ depend only on the responsiveness λ of population members. However, then the outcome of the negotiation is the same as if the mutant had just used $\tilde{r}(u, q)$ as a response rule in the negotiation, and the mutant can do no better than to adopt this simple response rule.

The function $\hat{r}_m = \hat{r}_f$ of Houston and Davies⁸ gives the best effort of an individual for a fixed effort of its partner. If we interpret this function as a negotiation rule (the HD rule) then the responsiveness under this rule is $\tilde{\lambda} = h(0)$ (as the partner's effort is fixed and hence has responsiveness $\lambda = 0$).

Within the class of response rules of the form of equation (6), the unique evolutionarily stable negotiation rule has responsiveness λ_{ESS} , which satisfies $\lambda_{\text{ESS}} = h(\lambda_{\text{ESS}})$. The other parameters are then given by $\mu_{\text{ESS}} = g(\lambda_{\text{ESS}})$ and $\rho_{\text{ESS}} = f(\lambda_{\text{ESS}})$. The stability of the outcome of negotiation is shown in Fig. 2. In Box 1 we show that

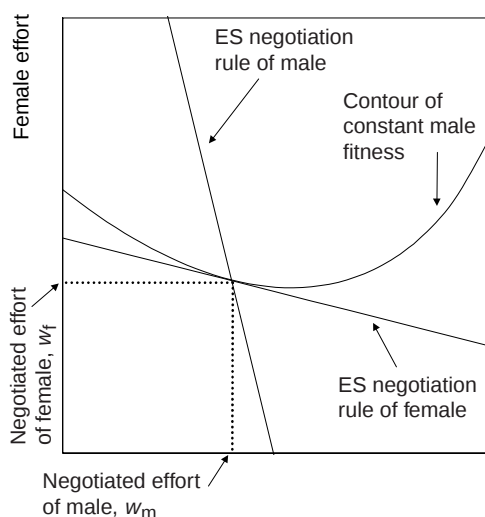


Figure 2 The evolutionarily stable negotiation rule for a male and its partner. If both players use this rule the resultant efforts will be w_m and w_f . The contour again shows constant fitness for the male. However, unlike in Fig. 1, this contour is now tangential to the evolutionarily stable negotiation rule of the female at (w_m, w_f) . Now any movement away from w_m will result in the male obtaining a lower fitness than if it had used the evolutionarily stable negotiation rule.

Box 1

Evolutionary stability

The evolutionarily stable (ES) effort of each parent, u^* , as predicted by the model of Houston and Davies⁸, is given by the solution to

$$\frac{\partial K}{\partial u}(u^*, 0) = B'(2u^*) \quad (8)$$

We assume that the marginal benefit to the brood decreases as parental effort increases. Thus B is a positive, increasing function and B'' is negative. Let $K_0 = \frac{\partial^2 K}{\partial u^2}(u^*, 0)$ and assume that the marginal cost of care increases as parental care increases, hence K_0 is positive. Let $K_1 = \frac{\partial^2 K}{\partial u \partial q}(u^*, 0)$, and assume that the marginal cost of care for a good quality individual is less than that for a poor quality individual, hence K_1 is negative.

The payoff to an individual of quality q providing effort u_1 whose partner provides effort u_2 is $R(u_1, u_2) = B(u_1 + u_2) - K(u_1, q)$. Taking the Taylor expansions (equations (4) and (5) to be exact and substituting from equation (8) we obtain $R(u_1, u_2) = (u_2 - u^*)B' + (u_1 + u_2 - 2u^*)^2 \frac{B''}{2} - (u_1 - u^*)^2 \frac{K_0}{2} - q(u_1 - u^*)K_1 - H(q)$, where $H(q)$ depends on q alone and B' and B'' are evaluated at $2u^*$.

We consider that females and males are identical and ignore any role asymmetries. We therefore refer to individuals' roles as player 1 or 2 rather than female or male. Consider a population in which an individual of quality q uses the response rule $r(u, q) = u^* + \rho + \mu q - \lambda(u - u^*)$, which is of the form of equation (6). We look for the optimal behaviour of a rare mutant within this population. Suppose that a rare mutant (player 1) of quality q_1 partners a resident population member (player 2) of quality q_2 . We allow the mutant to have full knowledge of its partner's quality and rule. We also allow the mutant to use any negotiating strategy providing that the negotiation eventually ends with well-defined efforts by both partners. If, after negotiation, player 1 provides effort u_1 , its payoff is

$$R(u_1, r(u_1, q_2)) = (r(u_1, q_2) - u^*)B' + (u_1 + r(u_1, q_2) - 2u^*)^2 \frac{B''}{2} - (u_1 - u^*)^2 \frac{K_0}{2} - q_1(u_1 - u^*)K_1 - H(q_1).$$

This is maximized when the final effort of player 1, $\hat{u}_1$, satisfies $\frac{\partial R}{\partial u_1}(\hat{u}_1, r(\hat{u}_1, q_2)) = 0$, that is $0 = r_u(\hat{u}_1, q_2)B' + 2(\hat{u}_1 + r(\hat{u}_1, q_2) - 2u^*) \times (1 + r_u(\hat{u}_1, q_2)) \frac{B''}{2} - 2(\hat{u}_1 - u^*) \frac{K_0}{2} - q_1 K_1$, where r_u denotes the partial derivative of r with respect to effort u . As $r_u(u_1, q_2) = -\lambda$, the mutant can do no better than negotiate a final effort $\hat{u}_1$ satisfying

$$0 = -\lambda B' + 2(\hat{u}_1 + r(\hat{u}_1, q_2) - 2u^*)(1 - \lambda) \frac{B''}{2} - 2(\hat{u}_1 - u^*) \frac{K_0}{2} - q_1 K_1 \quad (9)$$

Given that player 1 negotiates effort $\hat{u}_1$, the effort of player 2 after negotiation is $\hat{u}_2 = r(\hat{u}_1, q_1)$. Thus by equation (9), $\hat{u}_1$ and $\hat{u}_2$ are related by

$$0 = -\lambda B' + 2(\hat{u}_1 + \hat{u}_2 - 2u^*)(1 - \lambda) \frac{B''}{2} - (\hat{u}_1 - u^*)K_0 - q_1 K_1.$$

Rearranging this expression gives

$$\hat{u}_1 = u^* + \frac{-(\lambda B' + q_1 K_1 - (1 - \lambda)(\hat{u}_2 - u^*)B'')}{K_0 - (1 - \lambda)B''}.$$

Now consider a mutant rule under which the dependence of the negotiated effort of the mutant on its own quality q and its partner's negotiated effort u is given by

$$\tilde{r}(u, q) = u^* + \frac{-(\lambda B' + q K_1 - (1 - \lambda)(u - u^*)B'')}{K_0 - (1 - \lambda)B''} \quad (10)$$

Then the above analysis shows that, regardless of its quality, the mutant is doing the best it possibly can against every possible partner.

Given that a population uses the response rule $r(u, q) = u^* + \rho + \mu q - \lambda(u - u^*)$, a mutant that uses the response rule $\tilde{r}(u, q) = u^* + f(\lambda) + g(\lambda)q - h(\lambda)(u - u^*)$, where $f(\lambda) = \frac{-\lambda B'}{K_0 - (1 - \lambda)B''}$, $g(\lambda) = \frac{-K_1}{K_0 - (1 - \lambda)B''}$ and $h(\lambda) = \frac{-(1 - \lambda)B''}{K_0 - (1 - \lambda)B''}$ achieves the outcome of equation (10). Thus the optimal negotiation rule of a mutant is also a response rule of the form of equation (6).

From the above analysis there exists an ES negotiation rule within the class of response rules of the form of equation (6). This rule is given by $r_{\text{ESS}}(u, q) = u^* + \rho_{\text{ESS}} + \mu_{\text{ESS}}q - \lambda_{\text{ESS}}(u - u^*)$, where the responsiveness λ_{ESS} is a solution of the equation $\lambda_{\text{ESS}} = h(\lambda_{\text{ESS}})$, and the parameters μ_{ESS} and ρ_{ESS} are given by $\mu_{\text{ESS}} = g(\lambda_{\text{ESS}})$ and $\rho_{\text{ESS}} = f(\lambda_{\text{ESS}})$.

The responsiveness of the ES negotiation rule, λ_{ESS} , satisfies $\lambda_{\text{ESS}} = h(\lambda_{\text{ESS}})$ and thus is a solution of $B''\lambda_{\text{ESS}}^2 - (2B'' - K_0)\lambda_{\text{ESS}} + B'' = 0$. This equation has two roots; both are positive, only one is less than 1. We take λ_{ESS} to be the root satisfying $0 < \lambda_{\text{ESS}} < 1$. The parameter $\mu_{\text{ESS}} = g(\lambda_{\text{ESS}}) = \frac{-K_1}{K_0 - (1 - \lambda_{\text{ESS}})B''}$ is positive, as $K_0 > 0$, $K_1 < 0$, $B'' < 0$ and $1 - \lambda_{\text{ESS}} > 0$. The parameter $\rho_{\text{ESS}} = f(\lambda_{\text{ESS}}) = \frac{-\lambda_{\text{ESS}}B'}{K_0 - (1 - \lambda_{\text{ESS}})B''}$ is negative, as $B' > 0$ and $\lambda_{\text{ESS}} > 0$.

To show that the responsiveness of the ES negotiation rule is less than that of the HD rule we note that the function h is positive and decreasing and that λ_{ESS} is positive. Therefore $\lambda_{\text{ESS}} = h(\lambda_{\text{ESS}}) < h(0) = \hat{\lambda}$, the responsiveness of the HD rule.

The negotiated effort of a player of quality q_1 whose partner is quality q_2 is

$$u_{\text{ESS}}(q_1, q_2) = u^* + \frac{(1 - \lambda_{\text{ESS}})\rho_{\text{ESS}} + \mu_{\text{ESS}}(q_1 - \lambda_{\text{ESS}}q_2)}{(1 - \lambda_{\text{ESS}}^2)} \quad (11)$$

if both players use the ES negotiation rule, and $\hat{u}(q_1, q_2) = u^* + \frac{\hat{u}(q_1 - \lambda q_2)}{(1 - \lambda)}$ if both players use the HD rule.

It can be shown algebraically that so long as $q_1 + q_2$ is small enough that $q_1 + q_2 < \frac{(K_0 - 2B''B')}{B''K_1}$, then $u_{\text{ESS}}(q_1, q_2) < \hat{u}(q_1, q_2)$. Hence, the negotiated effort using the ES negotiation rule is smaller than that using the HD rule.

The negotiated effort of a player is a strictly increasing function of its quality (for a given partner's quality), as differentiating equation (11) gives $\frac{\partial u_{\text{ESS}}(q_1, q_2)}{\partial q_1} = \frac{\mu_{\text{ESS}}}{(1 - \lambda_{\text{ESS}}^2)} > 0$.

$0 < \lambda_{\text{ESS}} < \hat{\lambda}$. Thus the evolutionarily stable negotiation rule is less responsive than the HD rule. We also show that, if any pair of individuals use the evolutionarily stable negotiation rule, the outcome is that each individual expends less effort than if the individuals used the HD rule. Thus, if individuals recognize each other's quality without negotiation, every member of every pair expends more effort than if quality is not directly observable and effort has to be negotiated. In particular, if both pair members have quality $q = 0$, then the final negotiated effort of both is less than u^* . Moreover, as the outcomes under the evolutionarily stable negotiation rule do not depend on the degree of quality variation (provided there is some variation), we do not even obtain the Houston and Davies efforts in the limit as quality variation tends to zero.

Our analysis generalizes to any game in which individuals can respond to each other's actions, and shows that the action chosen as a result of negotiation may not be the best given that of the opponent. For example, when two animals inspect a potential predator^{12,13}, after negotiation the distance of one animal from the predator may not be the best distance given the distance of the other animal. In many signalling games, the receiver responds to the signaller but the signaller cannot respond to the behaviour of the receiver^{14–17}. A response by the signaller is, however, plausible in many situations. As an example, consider chicks using begging behaviour to signal need to a parent. If either foraging conditions for parents vary or all parents do not have the same foraging ability, then it will advantageous for a chick to alter its level of begging in response to the effort of its parent. This may result in negotiation

between the chick and the parent and the outcome may differ from that predicted by models in which there is no responding.

At evolutionary stability in our parental effort game, the negotiated effort of an individual's partner is a strictly increasing function of the partner's quality. Thus, after negotiation, an individual can infer the quality of its partner; that is, negotiated levels of effort are honest. Despite this, efforts are less than if individuals could directly observe quality. If population members behaved as if inferred quality had been directly observed, then a mutant individual that pretended to have lower quality in the negotiation phase would, after negotiation, expend low effort at the expense of its partner, who would compensate (Fig. 1). Thus, the population would not be evolutionarily stable. □

Received 2 March; accepted 1 July 1999.

1. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
2. Parker, G. A. & Maynard Smith, J. Optimality theory in evolutionary biology. *Nature* **348**, 27–33 (1990).
3. Pulliam, H. R., Pyke, G. H. & Caraco, T. The scanning behaviour of juncos: a game-theoretical approach. *J. Theor. Biol.* **95**, 89–103 (1982).
4. Lima, S. L. Vigilance while feeding and its relation to the risk of predation. *J. Theor. Biol.* **124**, 303–316 (1987).
5. McNamara, J. M. & Houston, A. I. Evolutionarily stable levels of vigilance as a function of group size. *Anim. Behav.* **43**, 641–658 (1992).
6. Wright, J. & Cuthill, I. C. Biparental care: short-term manipulation of partner contribution and brood size in the starling, *Sturnus vulgaris*. *Behav. Ecol.* **1**, 116–124 (1990).
7. Markman, S., Yom-Tov, Y. & Wright, J. Male parental care in the orange-tufted sunbird—behavioural adjustments in provisioning and nest guarding effort. *Anim. Behav.* **50**, 655–669 (1995).
8. Houston, A. I. & Davies, N. B. In *Behavioural Ecology: The Ecological Consequences of Adaptive Behaviour* (eds Sibly, R. M. & Smith, R. H.) 471–487 (Blackwell Scientific, Oxford, 1985).
9. Hatchwell, B. J. & Davies, N. B. Provisioning of nestlings by dunlocks, *Prunella modularis*, in pairs and trios: compensation reactions by males and females. *Behav. Ecol. Sociobiol.* **27**, 199–209 (1990).
10. Clutton-Brock, T. H. & Godfray, H. C. J. In *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 234–262 (Blackwell Scientific, Oxford, 1991).
11. Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, Oxford, 1997).
12. Dugatkin, L. A. *Cooperation Among Animals* (Oxford Univ. Press, Oxford, 1997).
13. Parker, G. A. & Milinski, M. Cooperation under predation risk: a data-based ESS analysis. *Proc. R. Soc. Lond. B* **264**, 1239–1247 (1997).
14. Grafen, A. Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546 (1990).
15. Maynard Smith, J. Honest signalling: the Philip Sidney game. *Anim. Behav.* **42**, 1034–1035 (1991).
16. Johnstone, R. A. Honest signalling, perceptual error and the evolution of 'all-or-nothing' displays. *Proc. R. Soc. Lond. B* **256**, 169–175 (1994).
17. Godfray, H. C. J. Signalling of need by offspring to their parents. *Nature* **352**, 328–330 (1995).

Acknowledgements

We thank I. Cuthill, S. Hunt, J. Hutchinson, I. Jewitt, P. Sozou and M. Yallop for comments on previous versions of this manuscript.

Correspondence and requests for materials should be addressed to J.M.M. (e-mail: john.mcnamara@bristol.ac.uk).

The liprin protein SYD-2 regulates the differentiation of presynaptic termini in *C. elegans*

Mei Zhen & Yishi Jin

Department of Biology, Sinsheimer Laboratories, University of California, Santa Cruz, California 95064, USA

At synaptic junctions, specialized subcellular structures occur in both pre- and postsynaptic cells. Most presynaptic termini contain electron-dense membrane structures¹, often referred to as active zones, which function in vesicle docking and release². The components of those active zones and how they are formed are largely unknown. We report here that a mutation in the *Caenorhabditis elegans* *syd-2* (for synapse-defective) gene causes a diffused localization of several presynaptic proteins and of a synaptic-vesicle membrane associated green fluorescent protein (GFP) marker^{3,4}. Ultrastructural analysis revealed that the active

zones of *syd-2* mutants were significantly lengthened, whereas the total number of vesicles per synapse and the number of vesicles at the prominent active zones were comparable to those in wild-type animals. Synaptic transmission is partially impaired in *syd-2* mutants. *syd-2* encodes a member of the liprin (for LAR-interacting protein) family of proteins which interact with LAR-type (for leukocyte common antigen related) receptor proteins with tyrosine phosphatase activity (RPTPs)^{5,6}. SYD-2 protein is localized at presynaptic termini independently of the presence of vesicles, and functions cell autonomously. We propose that SYD-2 regulates the differentiation of presynaptic termini in particular the formation of the active zone, by acting as an intracellular anchor for RPTP signalling at synaptic junctions.

We isolated the *syd-2(ju37)* mutant in a genetic screen for mutations affecting the localization of a GFP marker, *P_{unc-25}*—SNB-1::GFP, associated with the synaptic-vesicle membrane (refs 3, 4 and M.Z. and Y.J., unpublished). SNB-1 is a *C. elegans* synaptobrevin⁷, a synaptic-vesicle membrane protein involved in vesicle docking and exocytosis², and SNB-1::GFP is associated with vesicles³. The *unc-25* promoter drives the expression of SNB-1::GFP in the GABAergic DD and VD motorneurons^{4,8}. In wild-type animals carrying this marker, here called *juIs1*, GFP is seen as uniformly shaped fluorescent puncta evenly distributed along the ventral and dorsal nerve cords, which correspond to the presynaptic termini of 13 VD and 6 DD neurons, respectively⁹ (Fig. 1a, b). We found that, in *syd-2(ju37)* *juIs1* mutants, the total number of fluorescent puncta was the same as in wild-type animals (Fig. 1c–f). However, these puncta showed various degrees of diffusion, and the overall fluorescent intensity of GFP was reduced. In addition, three presynaptic proteins (synaptotagmin, syntaxin and RAB-3) were also diffusely localized in *syd-2* mutants (Fig. 1g, h; data not shown for syntaxin and RAB-3). *syd-2(ju37)* is a recessive mutation, and homozygous *syd-2(ju37)* animals showed mild defects in many visible behaviours, for example sluggish locomotion and defective egg laying. Immunocytochemistry and GFP markers revealed no abnormalities in the gross morphology of DD, VD and other motorneurons in *syd-2* mutants (data not shown). Thus, the diffuse SNB-1::GFP pattern in *syd-2(ju37)* mutants reflects a defect in presynaptic protein localization, but not in axonal guidance.

We cloned *syd-2* by germline transformation rescue (see Fig. 2 and Methods). The predicted SYD-2 protein is highly homologous to liprin- α (Fig. 2b)⁶. Liprin- α was first identified because it interacted with the second phosphatase domain of LAR-type RPTPs in the yeast two-hybrid assay⁵. In mammalian cells, liprin- α causes RPTPs to cluster to focal adhesions⁵. All liprins have coiled-coil domains at the amino terminal and three SAM domains at the carboxy terminal^{5,10}. SAM domains are protein modules that mediate homo- and heterodimerizations¹⁰. RPTPs interact with the C-terminal portion of liprin- α (ref. 5), and within this region SYD-2 and liprin- α 1 are 60% identical (Fig. 2b). In the yeast two-hybrid assay SYD-2 interacts with mammalian LAR and *Drosophila* DLAR (C. Serra-Pagès, M.Z., Y.J. and M. Streuli, unpublished results), indicating that SYD-2 is a functional homologue of liprin- α . Deletion of the three SAM domains (pCZ#10) greatly reduced *syd-2* rescuing ability (Fig. 2a), supporting the idea that the SAM domains are important for SYD-2 function. Moreover, the *ju37* mutation is probably a molecular null because it contained a C-to-T transition, changing glutamine 397 to an amber stop, and the predicted truncated protein was below the level of detection (Fig. 2b, c).

To gain insight into SYD-2 function, we analysed its expression pattern. Immunofluorescent staining of wild-type animals with antibodies against SYD-2 revealed distinct subcellular expression patterns in neurons and muscles (Fig. 3a–d). Using a *syd-2*–GFP reporter construct that expressed GFP in cell bodies (pCZ#14; Fig. 2a), we determined that *syd-2* was expressed in all neurons and muscles. Two lines of evidence indicate that *syd-2* is required in

neurons. First, we performed genetic mosaic analysis using *syd-2(ju37) juIs1* animals carrying an extrachromosomal array containing a *syd-2(+)* genomic clone and a marker, SUR-5::GFP, that is expressed in the nuclei of all somatic cells¹¹ (see Methods). Animals that had lost *syd-2(+)* in the AB lineage (which generates all motor neurons) but not the P₁ lineage (which generates all but one of the body-muscle cells) showed diffuse localization of the SNB-1::GFP

marker. These AB-loss mosaic animals were as uncoordinated and defective in egg-laying as *syd-2* mutants. In contrast, animals lacking *syd-2(+)* in the P₁ but not the AB lineage had normal SNB-1::GFP localization, and were behaviourally wild type. This result indicated that SYD-2 is required in neurons for normal SNB-1::GFP localization. Second, we expressed SYD-2 in GABAergic motor neurons of *syd-2(ju37)* mutants using the *unc-25* promoter⁸. Antibody staining

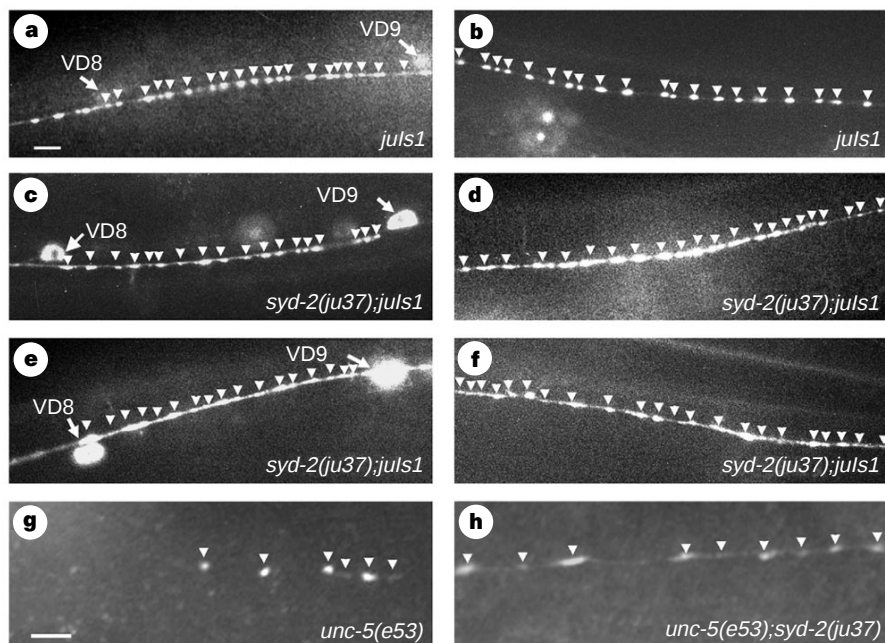


Figure 1 *syd-2(ju37)* causes diffuse localization of a synaptic-vesicle GFP marker. **a, b**, Expression of *P_{unc-25}*-SNB-1::GFP (*juIs1*) in a wild-type animal. Shown are the presynaptic regions of VD9 (**a**) and DD5 (**b**). Arrowheads, SNB-1::GFP fluorescent puncta, which were 0.2–0.5 μ m in diameter and $\sim$ 3 μ m apart; arrows, VD cell bodies. **c–f**, *juIs1* expression in a *syd-2(ju37)* animal at the same regions as in **a** (**c**, **e**) and **b** (**d**, **f**). SNB-1::GFP puncta were moderately (**c**, **d**) or severely (**e**, **f**) diffused.

g, h, Endogenous synaptotagmin (SNT-1) expression in *unc-5(e53)* (**g**) and *unc-5(e53);syd-2(ju37)* (**h**) animals. In *unc-5(e53)* animals DDs form ectopic synapses that have normal ultrastructural morphology²⁸, allowing examination of SNT-1 localization at individual synapses. The fluorescent puncta (arrowheads) in **h** were more diffuse than those in **g**, but the total fluorescent intensity was comparable. Scale bars: **a–f**, 5 μ m; **g, h**, 8 μ m.

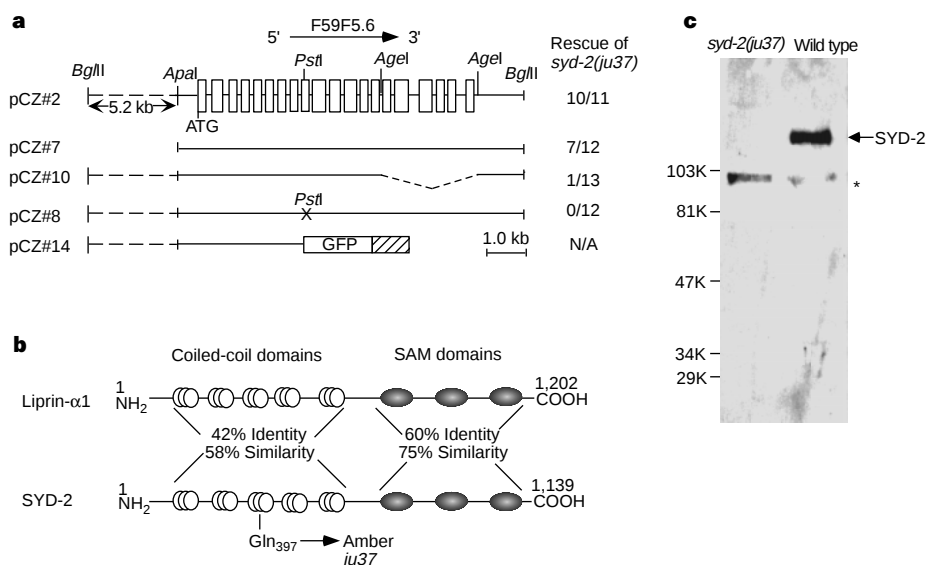


Figure 2 *syd-2* encodes a liprin. **a**, Rescue of *syd-2(ju37)*. Numbers on the right indicate the number of rescued transgenic lines out of the total number of lines scored. Open boxes, exons; x, a frameshift mutation at the *PstI* site. **b**, Protein comparison of SYD-2 and mammalian liprin α 1 (ref. 6). SYD-2 has coiled-coil domains at amino acids 31–140, 178–297, 331–432, 457–515 and 655–699, and SAM domains at amino acids 876–

940, 961–1,023 and 1,046–1,116. **c**, Western blot of wild-type and *syd-2(ju37)* worm lysates with anti-SYD-2 antibodies. The full-length SYD-2 protein is 130K. The resulting truncated 43K protein in *syd-2(ju37)* was below detection level. * crossreacting protein (see Methods).

Table 1 Synaptic-vesicle number and active-zone length in *syd-2(ju37)* and wild-type NMJs

Type of NMJs	Genotype	Average no. of synaptic vesicles (<i>n</i>)			Length of active zones (nm) (<i>n</i>)
		Total per synapse	In mid-synaptic region	Total in sections with active zones	
GABAergic	Wild type	141.0 ± 12.8* (28)	30.4 ± 2.6* (32)	56.4 ± 7.6* (28)	101 ± 8* (32)
	<i>syd-2(ju37)</i>	161.8 ± 18.4 (13)	37.5 ± 4.7 (22)	103.1 ± 16.8 (15)	207 ± 13 (22)
Cholinergic	Wild type	150.4 ± 12.7 (10)	33.0 ± 2.4 (11)	67.9 ± 7.6 (11)	132 ± 11 (15)
	<i>syd-2(ju37)</i>	162.7 ± 21.8 (7)	31.9 ± 3.8 (10)	103.0 ± 12.0 (10)	235 ± 17 (12)

Total synaptic-vesicle number per synapse is the sum of synaptic vesicles in all sections containing active zones plus flanking two sections. The mid-synaptic region is the section with the most prominent active-zone structure in individual synapses. *n*, number of synapses examined. Length of active zone was calculated by multiplying the total number of serial sections containing active zones by 60 nm. * s.e.m.

detected SYD-2 only in these neurons, and the SNB-1::GFP defects in the neurons were completely rescued. Taken together, we conclude that SYD-2 functions cell autonomously in presynaptic neurons.

To determine where SYD-2 is localized in presynaptic neurons, we analysed the distribution of SYD-2 with respect to the SNB-1::GFP marker. As SYD-2 is widely expressed in neurons and muscles, we stained *syd-2(ju37) juIs1* animals carrying a *P_{unc-25}-syd-2(+)* transgene using antibodies against SYD-2 and GFP (for SNB-1::GFP). Each fluorescent punctum of SNB-1::GFP was closely associated with a smaller cluster of SYD-2 protein (Fig. 3e–g), with SYD-2 being surrounded by SNB-1::GFP (Fig. 3h, i). This localization of SYD-2 protein was not altered in *unc-104* kinesin mutants (data not shown), in which synaptic vesicles are retained in the cell bodies¹². Thus, SYD-2 is localized to the presynaptic region, but is not a component of synaptic vesicles.

The diffuse appearance of SNB-1::GFP in *syd-2* mutants could be due to altered distribution of synaptic vesicles, overproduction of synaptic vesicles or defects in the structure of the presynaptic termini. To distinguish between these possibilities, we examined the chemical synapses in *syd-2(ju37)* mutants using electron microscopy (see Methods). These mutants had the same number of chemical synapses as had wild-type animals. The total number of synaptic vesicles per synapse in *syd-2* animals was on average comparable to that of wild-type animals (Table 1). Moreover, the number of vesicles in the mid-synaptic region at individual synapses

was not significantly different in wild-type and *syd-2(ju37)* animals (Table 1). The electron-dense active zones, however, were significantly lengthened in the mutant animals (Fig. 4, Table 1). The width of the active zones was not significantly altered, consistent with our observation that the SNB-1::GFP marker was diffused only along the length of axons in *syd-2(ju37)* animals (Fig. 1c–f). The increase in the active-zone length in these animals was correlated with an increase of the sum of synaptic vesicles found in all sections containing active zones for individual synapses (Table 1). We noted that most of the active-zone lengthening in *syd-2(ju37)* animals occurred in regions flanking the prominent active zone, and that the prominent active zones were consistently less electron-dense in *syd-2(ju37)* animals than were those in wild-type animals. Our data thus show that the diffusion of SNB-1::GFP in *syd-2* mutants does not reflect alterations in vesicle production and localization.

We used pharmacological assays to investigate how the increased size of active zones in *syd-2(ju37)* mutants affects synaptic transmission. *syd-2* animals retained eggs moderately: on average, 25 eggs were retained in an adult hermaphrodite compared with an average of 12 eggs in a wild-type adult of the same age (*n* = 40). Exogenous serotonin and imipramine stimulate egg laying, presumably by mimicking and enhancing endogenous serotonin released by the HSN neurons¹³, which innervate the sex muscles. *syd-2* mutant animals laid eggs when treated with serotonin but not with imipramine (Fig. 5a), consistent with a presynaptic defect in

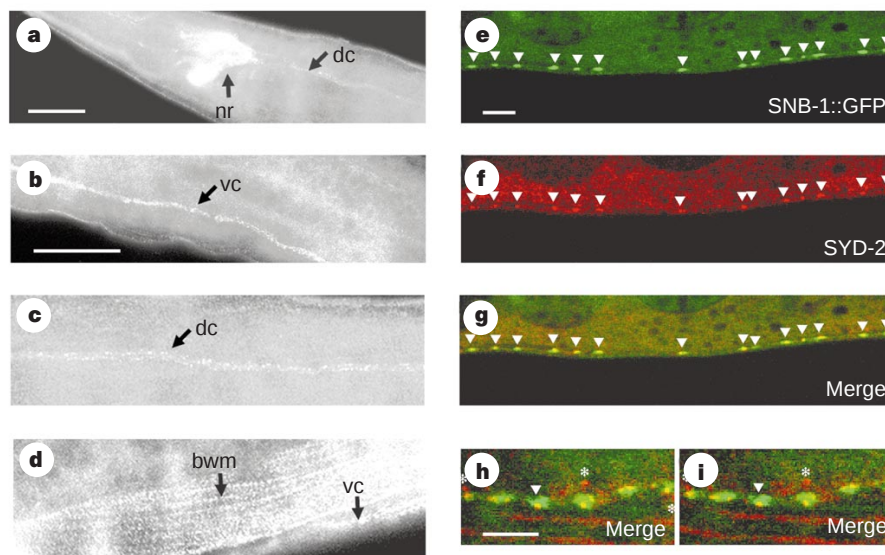


Figure 3 SYD-2 is localized to presynaptic regions. **a–d**, SYD-2 expression in a wild-type animal detected by immunocytochemistry with anti-SYD-2 antibodies, nr, nerve ring; vc, ventral cord; dc, dorsal cord; bwm, body-wall muscle. SYD-2 was localized between dense bodies in muscles (not shown). In *syd-2(ju37)*, neuronal expression was eliminated and muscle expression was greatly reduced (not shown). **e–i**, Confocal images of subcellular localization of SYD-2 in a VD neuron. **e–g**, lateral views of the presynaptic

termini, showing that SYD-2 and SNB-1::GFP were associated at a 1:1 ratio (arrowheads). **h, i**, Top views, showing that SYD-2 was surrounded by SNB-1::GFP. The focal plane in **i** was 0.15 μ m more dorsal than that in **h**; SYD-2 in **i** was more separated from SNB-1::GFP than in **h**. * Perinuclear SYD-2, presumably caused by overexpression. Scale bars: **a–d**, 50 μ m; **e–h**, 5 μ m.

HSNs. Aldicarb, an inhibitor of cholinesterase, causes wild-type worms to hypercontract and die⁷. Mutants defective in synaptic transmission, such as *snt-1*, *rab-3* and *aex-3*, are resistant to aldicarb to various degrees⁷, and *syd-2* mutants were resistant to aldicarb to a similar degree to *rab-3* and *aex-3* mutants (Fig. 5b). This analysis indicates that synaptic transmission was partially impaired in *syd-2* mutants.

Presynaptic termini are specialized both morphologically and molecularly to accommodate synaptic transmission¹⁴, but the formation of such specialized structures is poorly understood. Most of the known presynaptic proteins function in vesicle exocytosis and endocytosis². Only a few presynaptic proteins have been implicated in presynaptic structure formation, including the *Drosophila* Discs-Large¹⁵, Fasciclin II (ref. 15) and Still-life¹⁶, and the vertebrate synapsins¹⁷, Rim¹⁸ and Bassoon¹⁹. Discs-Large and Fasciclin II regulate the number of active zones per synapse¹⁵, whereas synapsins seem to mediate vesicle recruitment or maturation at the active zone¹⁷. How the other proteins affect presynaptic structures is not clear. We show here that the liprin protein SYD-2 acts within presynaptic neurons. In *syd-2* mutants a presynaptic marker and presynaptic proteins are diffusely localized, and the presynaptic active zones are lengthened, but the total vesicle number per synapse

is not significantly altered. In contrast, *C. elegans* mutants in which the number of synaptic vesicles is either reduced (as in *snt-1* and *rab-3* mutants^{20,21}) or increased (for example, *unc-13*; E. Jorgensen, personal communication) have normal active zones, indicating that the size of the active zones does not strictly depend on vesicle number.

What is the role of *syd-2* in synapses? The protein homology of SYD-2 suggests that it interacts with RPTPs. Interestingly, the inhibition of serine/threonine phosphatases by okadaic acid results in inefficient vesicle clustering^{22,23}, but it is not known whether the inhibition of RPTPs has similar effects. The SNB-1::GFP phenotype in *syd-2* mutants could be explained either by an increased plasma-membrane area for vesicle exocytosis and endocytosis or by inefficient vesicle clustering. The SNB-1::GFP marker is localized normally in *rab-3* and *unc-13* mutants (E. Jorgensen, personal communication; and our unpublished data), but is diffused in *snt-1* mutants, which have defective endocytosis²⁰. Because the number of vesicles present in the prominent active zones in *syd-2* mutants is similar to that in wild type (Table 1), we favour the interpretation that the diffused SNB-1::GFP represents an altered association of this marker with the plasma membrane in *syd-2* mutants, rather than altered vesicle distribution. However, our data do not exclude the possibility that subtle changes in vesicle distribution occur in *syd-2* mutants, because, for technical reasons, we could examine only a limited number of synapses by electron microscopy. *syd-2* mutants show an increase in the total number of vesicles clustered in regions containing active zones (Table 1), yet synaptic transmission is impaired, suggesting that the active zones in *syd-2* are defective. The structural and functional defects of *syd-2* mutants are mild, which suggests that *syd-2* may function redundantly with other proteins. We propose that SYD-2 may act as a structural component of the active zones and/or as a cytoplasmic anchor to recruit other molecules to the active zones. This is

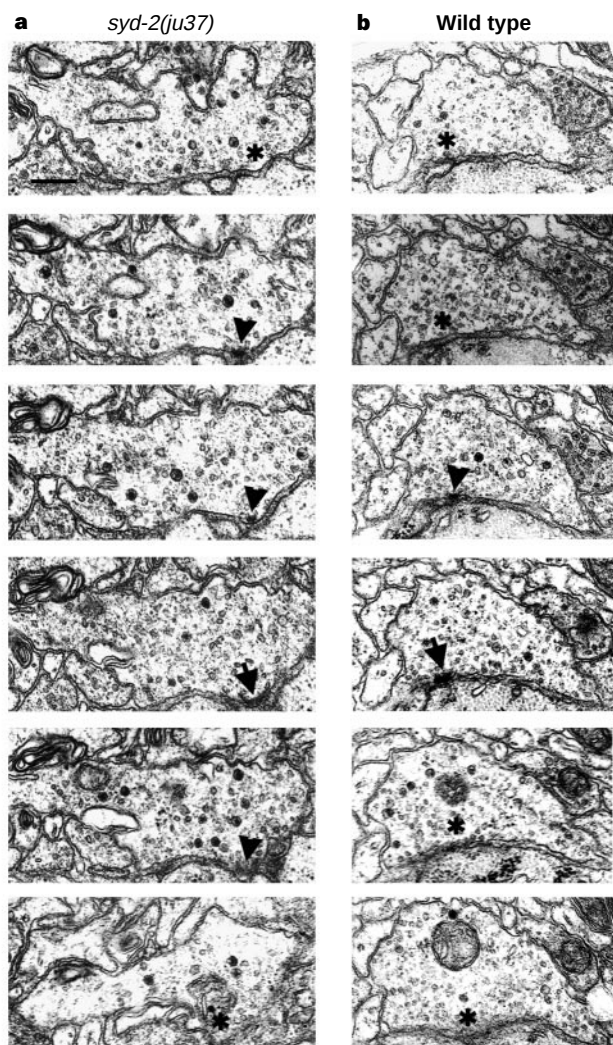


Figure 4 Presynaptic active zones were lengthened in *syd-2(ju37)* animals. **a**, Serial electron micrographs (EMs) of a GABAergic NMJ made by the VD2 motor neuron in *syd-2(ju37)* animals. **b**, Serial EM sections of a GABAergic NMJ made by VD3 in wild-type animals. Arrows, prominent active zones; arrowheads, smaller active-zone structures; * No active zones. Scale bar, 200 nm.

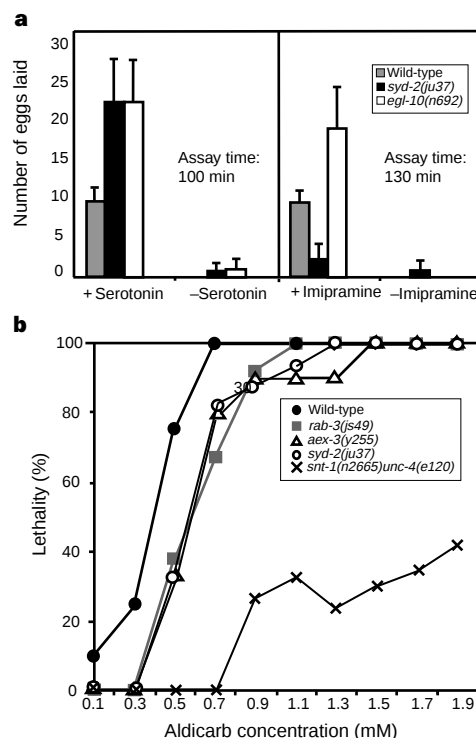


Figure 5 Pharmacological responses of *syd-2(ju37)* animals. **a**, Serotonin but not imipramine stimulated *syd-2(ju37)* to lay eggs. *egl-10(n692)* (ref. 13) was used as a control. *syd-2(ju37)* and *egl-10(n692)* animals laid more eggs than did wild-type (N2) animals in both assays because they normally retained more eggs than did wild-type animals. Error bars, standard deviations ($n = 12$). **b**, *syd-2(ju37)* was moderately resistant to aldicarb, comparable to *rab-3(js49)* and *aex-3(y255)*.

consistent with the observation that mammalian liprins recruit RPTPs to focal adhesions, and may thus regulate the disassembly of focal adhesions⁵. The roles of RPTPs in synapse formation have not been examined. In *Drosophila*, RPTPs are required for axon defasciculation²⁴, a process that may also involve disassembly of protein complexes mediating cell–cell interactions at membranes. *syd-2* is the first liprin identified by genetics and will provide an *in vivo* system for studying the signalling pathway in which liprins, and possibly RPTPs, are involved at synaptic junctions. □

Methods

C. elegans genetics.

Worms were grown on NGM plates at 22.5 °C as described²⁵. *syd-2(ju37)* was isolated in strain CZ323 *sem-4(n1378); juIs1* after ethyl methane sulphonate mutagenesis (M.Z. and Y.J., unpublished). *syd-2(ju37)* was mapped to linkage group X based on the diffuse SNB-1::GFP phenotype, and was further localized between *unc-115(e2225)* and *egl-15(n484)*, in the vicinity of *vab-3(e648)*. From *unc-115(e2225) egl-15(n484)/syd-2(ju37)* heterozygotes, two out of four *Unc* non-*Egl* and three out of four *Egl* non-*Unc* recombinants segregated *syd-2*. From *unc-115(e2225)vab-3(e648)/syd-2(ju37)* heterozygotes, all four *Unc* non-*Vab* and none of three *Vab* non-*Unc* animals segregated *syd-2*.

syd-2 cloning.

Four cosmids from the *vab-3* region were injected into *syd-2(ju37)* mutants, and cosmid F59F5 was found to contain *syd-2* rescuing activity. A 12.7-kilobase (kb) *Bgl*II fragment from F59F5, containing F59F5.6, was cloned into pBluescript SK (Stratagene) to generate the clone pCZ#2. All other *syd-2* clones in Fig. 2a were generated from pCZ#2 following standard protocols. pCZ#14 was constructed using the GFP vectors provided by A. Fire (personal communication).

Germline transformation was performed using pRF4 as a co-injection marker as described²⁶. Plasmids and cosmids were injected at 50–100 ng μ l⁻¹ and 20–30 ng μ l⁻¹, respectively. Integrated stable lines were established from transgenic animals by X-ray-induced mutagenesis.

Mosaic analysis of syd-2.

CZ1123 *syd-2(ju37)juIs1; juEx169[syd-2(+); SUR-5::GFP]* was generated by co-injecting *syd-2(ju37)juIs1* animals with pCZ#2, a plasmid containing the wild-type *syd-2* genomic DNA, and pTG96, a plasmid containing the *SUR-5::GFP* DNA that expresses GFP in all somatic nuclei¹¹. Animals that lost the *juEx169[syd-2(+); SUR-5::GFP]* extrachromosomal array in P₁ lineage retained *SUR-5::GFP* expression in neurons and hypodermis, but not in gut cells and body wall and vulval muscles. Animals that lost the array in the AB lineage expressed *SUR-5::GFP* in the nuclei of muscle and gut cells, but not in neurons and hypodermis. Five P₁-loss mosaic and six AB-loss mosaic animals were examined for *juIs1* expression pattern, locomotion and egg laying.

SYD-2 antibody production and immunocytochemistry.

A His₆-SYD-2 fusion protein containing amino acids 206–704 (from the *yk100c11* complementary DNA clone) was expressed in bacteria using the pRSETB vector (Qiagen) following the manufacturer's protocol, and used to immunize rabbits (Animal Pharm). Antiserum was affinity purified using an Affi-gel 15 column coupled with purified His₆-SYD-2 following the manufacturer's instructions (Bio-Rad). Worm protein extracts were made in 3% SDS, 10% glycerol and 3% β -mercaptoethanol. On western blots the affinity-purified SYD-2 antibodies detected a protein of relative molecular mass 130,000 (130K) which corresponds to the predicted size for full-length SYD-2 and was present only in wild-type worms. The anti-SYD-2 antibodies also detected a 96K protein that was present in both wild-type and *syd-2(ju37)* worms (Fig. 2c). We concluded that this 96K protein was not from the *syd-2* locus for the following reasons: the amount of this protein was unchanged in lysates of animals that overexpressed *syd-2*; the size of the protein was unchanged in lysates of animals carrying a rescuing GFP-tagged full-length SYD-2 transgene, whereas the 130K protein was shifted to the expected higher molecular weight; and no alternative RNA splice forms were detected from *syd-2* by RT-PCR.

Whole-mount immunofluorescent staining was performed as described²⁷. Anti-SYD-2 antibodies were used at 0.03 μ g μ l⁻¹ on western blots and 6 μ g μ l⁻¹ on whole mounts. Monoclonal anti-GFP antibodies (from S. Mitani) were used at 0.2 μ g μ l⁻¹. Cy5-conjugated goat anti-rabbit and FITC-conjugated goat anti-mouse IgG secondary antibodies (Cappel) were used at 0.375 μ g μ l⁻¹.

Drug tests.

Aldicarb, serotonin and imipramine tests were performed as described^{13,20}. All drug tests were scored blind and repeated at least three times.

Electron microscopy.

Wild-type (N2), CZ900 *syd-2(ju37)*, CZ458 *syd-2(ju37)juIs1* and CZ333 *juIs1* animals were fixed with glutaraldehyde and osmium tetroxide as described⁸. Fixed and embedded worms were first cut with glass knives at the posterior pharyngeal bulb. Starting from the anterior gonad, 200–400 serial thin sections of 60 nm thickness were collected for each sample. The ventral and/or dorsal nerve cords were photographed and examined. Chemical synapses were identified by the accumulation of synaptic vesicles around the

darkly stained membrane structures (active zones). GABAergic neuromuscular junctions (NMJs) were identified as synapses to muscle arms made by DD/VD neurons, which lie on the outer surface of the ventral cord. Cholinergic NMJs were identified as dyadic synapses to both DD or VD neurons and muscles⁴. Three N2 animals and two animals of each other genotype were examined. The data analysis was carried out by a double-blind test.

Received 24 March; accepted 19 July 1999.

- Landis, D. M. D., Hall, A. K., Weinstein, L. A. & Reese, T. S. The organization of cytoplasm at the presynaptic active zone of a central nervous system synapse. *Neuron* **1**, 201–209 (1988).
- Südhof, T. C. The synaptic vesicle cycle: a cascade of protein–protein interactions. *Nature* **375**, 645–653 (1995).
- Nonet, M. L. Visualization of synaptic specializations in live *C. elegans* with synaptic vesicle protein–GFP fusions. *J. Neurosci. Methods* **89**, 33–40 (1999).
- Hallam, S. & Jin, Y. *lin-14* regulates the timing of synaptic remodeling in *Caenorhabditis elegans*. *Nature* **395**, 78–80 (1998).
- Serra-Pages, C. *et al.* The LAR transmembrane protein tyrosine phosphatase and a coiled-coil LAR-interacting protein co-localize at focal adhesions. *EMBO J.* **14**, 2827–2838 (1995).
- Serra-Pages, C., Medley, Q. G., Tang, M., Hart, A. & Streuli, M. Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins. *J. Biol. Chem.* **273**, 15611–15620 (1998).
- Rand, J. B. & Nonet, M. Synaptic transmission. In *C. elegans II* (eds Riddle, D., Blumenthal, T., Meyer, B. & Priess, J. R.) 611–643 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Jin, Y., Jørgensen, E., Hartwig, E. & Horvitz, H. R. The *Caenorhabditis elegans* gene *unc-25* encodes glutamic acid decarboxylase and is required for synaptic transmission but not synaptic development. *J. Neurosci.* **19**, 539–548 (1999).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Stapleton, D., Balan, I., Pawson, T. & Sicheri, F. The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nature Struct. Biol.* **6**, 44–49 (1999).
- Yochem, J., Gu, T. & Han, M. A new marker for mosaic analysis in *Caenorhabditis elegans* indicates a fusion between hyp6 and hyp7, two major components of hypodermis. *Genetics* **149**, 1323–1334 (1998).
- Hall, H. & Hedgecock, M. Kinesin-related gene *unc-104* is required for axonal transport of synaptic vesicles in *C. elegans*. *Cell* **65**, 837–847 (1991).
- Trent, C., Tsung, N. & Horvitz, H. R. Egg-laying defective mutants of the nematode *C. elegans*. *Genetics* **104**, 619–647 (1983).
- Burns, M. E. & Augustine, G. J. Synaptic structure and function: dynamic organization yields architectural precision. *Cell* **83**, 187–194 (1995).
- Thomas, U. *et al.* Synaptic clustering of the cell adhesion molecule Fasciclin II by Discs-Large and its role in the regulation of presynaptic structure. *Neuron* **19**, 787–799 (1997).
- Sone, M. *et al.* Still life, a protein in synaptic terminals of *Drosophila* homologous to GDP-GTP exchangers. *Science* **275**, 543–547 (1998).
- Rosahl, T. W. *et al.* Essential functions of synapsins I and II in synaptic vesicle regulation. *Nature* **375**, 488–493 (1995).
- Wang, Y., Okamoto, M., Schmitz, F., Hofmann, K. & Südhof, T. C. Rim is a putative RAB3 effector in regulating synaptic-vesicle fusion. *Nature* **388**, 593–598 (1997).
- tom Dieck, S. *et al.* Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized at the active zone of presynaptic nerve termini. *J. Cell Biol.* **142**, 499–509 (1998).
- Jørgensen, E. *et al.* Defective recycling of synaptic vesicles in synaptotagmin mutant of *Caenorhabditis elegans*. *Nature* **378**, 196–199 (1995).
- Nonet, M. *et al.* *Caenorhabditis elegans rab-3* mutant synapses exhibit impaired function and are partially depleted of vesicles. *J. Neurosci.* **17**, 8061–8073 (1997).
- Betz, W. J. & Henkel, A. W. Okadaic acid disrupts clusters of synaptic vesicles in frog motor nerve terminals. *J. Cell Biol.* **124**, 843–854 (1994).
- Kraszewski, K. *et al.* Synaptic vesicle dynamics in living cultured hippocampal neurons visualized with CY3-conjugated antibodies directed against the luminal domain of synaptotagmin. *J. Neurosci.* **15**, 4328–4342 (1995).
- Desai, C. J., Gindhart, J. G. Jr, Goldstein, L. S. & Zinn, K. Receptor tyrosine phosphatases are required for motor axon guidance in the *Drosophila* embryo. *Cell* **84**, 599–609 (1996).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequence. *EMBO J.* **10**, 3959–3970 (1991).
- Finney, M. & Ruvkun, G. B. The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* **63**, 895–905 (1990).
- Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The *unc-5*, *unc-6*, *unc-40* genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **2**, 61–85 (1990).

Acknowledgements

We thank M. Nonet for SNB-1::GFP DNA and antibodies for SNT-1, SNB-1, syntaxin and RAB-3; Y. Kohara for the *yk100c11* clone; A. Fire for the GFP vectors; M. Han for the pTG96 plasmid; C. elegans genome consortium for the sequence and clone of cosmid F59F5; A. Chisholm, E. Jørgensen and M. Nonet for helpful discussion throughout this work; E. Hartwig for teaching us electron micrograph analysis; and R. Horvitz for his generosity. A. Chisholm, J. Kaplan, G. Garriga, L. Hinck, A. Zahler, R. Baran, Y. Ho and members of the Chisholm and Jin laboratories made comments on this paper. Some strains used in this work were obtained from the *Caenorhabditis elegans* Genetics Center, which is funded by NIH Center for research resources. M.Z. is supported by a Human Frontier Science Program Long Term Postdoctoral Fellowship. Y.J. is a Alfred P. Sloan research fellow. This work was supported by US PHS research grant NS35546 to Y.J. The GenBank accession numbers for *syd-2* genomic DNA and cDNA are Z50794 and AF170122, respectively.

Correspondence and requests for materials should be addressed to Y.J. (e-mail: jin@biology.ucsc.edu).

Familial dementia caused by polymerization of mutant neuroserpin

Richard L. Davis^{††}, Antony E. Shrimpton^{††}, Peter D. Holohan^{†‡}, Charles Bradshaw[§], David Feiglin^{||}, George H. Collins⁺, Peter Sonderegger[¶], Jochen Kinter[#], Lyn Marie Becker^{*}, Felicitas Lacbawan^{**}, Donna Krasnewich^{**}, Maximilian Muenke^{**}, Daniel A. Lawrence^{††}, Mark S. Yerby^{‡‡}, Cheng-Mei Shaw^{§§}, Bibek Goopu^{||||}, Peter R. Elliott^{||||}, John T. Finch^{##}, Robin W. Carrell^{||||} & David A. Lomas^{|||}

Departments of ^{*}Clinical Pathology, [‡]Pharmacology, [§]Neurology, ^{||}Radiology, State University of New York Health Science Center, Syracuse, New York 13210, USA

[¶]Department of Biochemistry, University of Zurich, Zurich CH-8057, Switzerland

^{**}National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

^{††}American Red Cross, Holland Laboratories, Rockville, Maryland 20855, USA

^{‡‡}Departments of Neurology, Public Health and Obstetrics-Gynecology, Oregon Health Sciences University, Portland, Oregon 97210, USA

^{§§}Department of Pathology, University of Washington, School of Medicine, Seattle, Washington 98104, USA

Departments of ^{||||}Haematology and ^{¶¶}Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge CB2 2XY, UK

^{##}Laboratory of Molecular Biology, MRC Centre, Cambridge CB2 2XY, UK

[†]These authors contributed equally to this work

Aberrant protein processing with tissue deposition is associated with many common neurodegenerative disorders^{1,2}; however, the complex interplay of genetic and environmental factors has made it difficult to decipher the sequence of events linking protein aggregation with clinical disease³. Substantial progress has been made toward understanding the pathophysiology of prototypical conformational diseases and protein polymerization in the superfamily of serine proteinase inhibitors (serpins)^{4,5}. Here we describe a new disease, familial encephalopathy with neuroserpin inclusion bodies, characterized clinically as an autosomal dominantly inherited dementia, histologically by unique neuronal inclusion bodies and biochemically by polymers of the neuron-specific serpin, neuroserpin^{6,7}. We report the cosegregation of point mutations in the neuroserpin gene (*PII2*) with the disease in two families. The significance of one mutation, S49P, is evident from its homology to a previously described serpin mutation⁸, whereas that of the other, S52R, is predicted by modelling of the serpin template. Our findings provide a molecular mechanism for a familial dementia and imply that inhibitors of protein polymerization may be effective therapies for this disorder and perhaps for other more common neurodegenerative diseases.

We have found familial encephalopathy with neuroserpin inclusion bodies (FENIB) in two Caucasian families living in the USA; both are affected by pre-senile dementia. In the larger family, affected individuals present clinically around the fifth decade of life with cognitive decline including deficits in attention and concentration, response regulation difficulties and impaired visuospatial skills. Memory is also impaired, but to a lesser degree than is typically seen in patients with Alzheimer's disease. After several years of disease progression, most affected individuals are unable to work and eventually require nursing-home care. The second, much smaller family shows an earlier clinical onset, during the second or third decades of life, and presents with both epilepsy and progressive cognitive decline ending in institutionalization⁹. The principal neuropathologic finding in both families is the presence of eosinophilic neuronal inclusions distributed throughout the deeper layers of the cerebral cortex and in many subcortical nuclei, especially the substantia nigra (Fig. 1a). These inclusions, which we have named Collins bodies, are round, 5–50 µm in diameter and strongly periodic acid-Schiff (PAS) positive, but diastase resistant. They are distinctly different from numerous previously described entities, including Lewy bodies, Pick bodies and Lafora bodies¹⁰. The inclusions were obtained as an enriched fraction from the postmortem brain of an affected individual (Fig. 1b), and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) showed them to be composed primarily of a single major protein (Fig. 1c), identified as neuroserpin from amino-acid sequencing (data not shown) and confirmed by western blotting (Fig. 1c). The neuroserpin protein consists of 410 amino acids and is glycosylated to a final gene product of relative molecular mass 55,000 (*M_r* 55K) (ref. 11). The neuroserpin gene (*PII2*) has been mapped to 3q26 (ref. 11). DNA sequencing of polymerase chain reaction (PCR) products designed to flank all eight coding exons of two affected individuals from the large family revealed only a single point mutation: a T to C transition at nucleotide 226 (Fig. 2a). This mutation, which we identify as *PII2*_{Syracuse}, results in the substitution of Ser49 in neuroserpin by proline (S49P). The T to C transition creates an *MnII* restriction site, which makes screening for this mutation in the family straightforward. Therefore, primers were designed to generate a 344-base-pair (bp) PCR product covering this selected region of exon 2. *MnII* digests of the PCR product generated from the wild-type neuroserpin gene give two fragments, of 254 and 90 bp. However, the T to C transition in *PII2*_{Syracuse} creates an *MnII* restriction site within the 254-bp product, giving additional products of 158 and 96 bp. Because affected individuals are heterozygotes, *MnII* restriction digests of the PCR product of exon 2 will give fragments of 254, 158, 96 and 90 bp for individuals carrying this mutation. This mutation occurs in 14 individuals out of the 38 family members screened (Fig. 2b, c); conversely, it was not detected upon screening 120 chromosomes from 60 unrelated Caucasian individuals.

The disease phenotype was ascertained either by observing

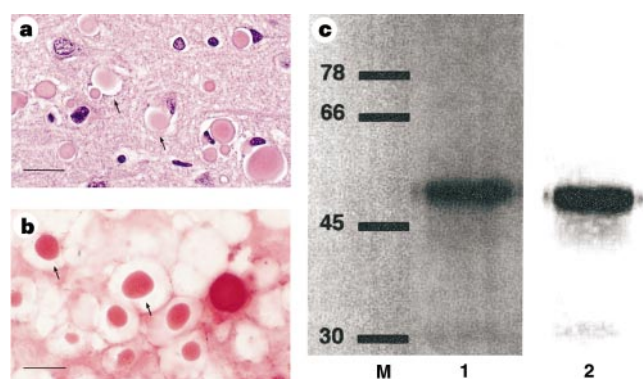
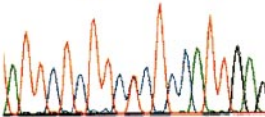


Figure 1 Identification, purification and composition of Collins bodies. **a**, Haematoxylin and eosin (H and E) stained section of cerebral cortex showing presence of Collins bodies (arrows). **b**, Purified fraction of Collins bodies stained with H and E. Scale bar, 50 µm. **c**, SDS-PAGE of purified Collins bodies solubilized by heating at 75 °C in 4% SDS for 2 h: lane 1, Coomassie blue stain; lane 2, western blot with antineuroserpin antibody.

a

ATTCTCTTC**N**CTCCATTGAG
210 220



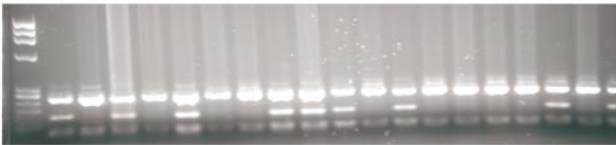
ATT CTC TTC **N**CT CCA TTG AGT
Ile Leu Phe Pro Leu Ser

ATT CTC TTC **T**CT CCA TTG AGT
Ile Leu Phe **Ser** Pro Leu Ser

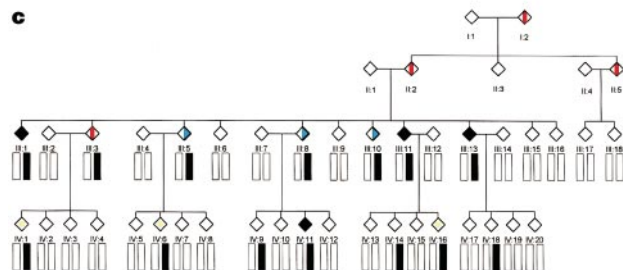
ATT CTC TTC **C**CT CCA TTG AGT
Ile Leu Phe **Pro** Pro Leu Ser

b

PhiX174/*Hae*III
marker III:1 III:2 III:3 III:4 III:5 III:6 III:7 III:8 III:9 III:10 III:11 III:12 III:13 III:14 III:15 III:16 III:17 IV:1 IV:2 IV:3

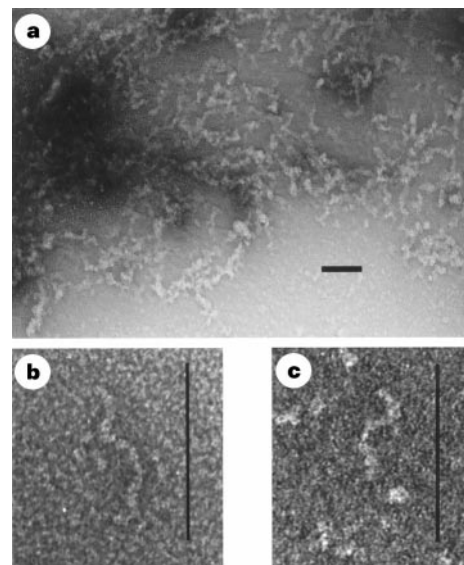


M M M M M M M M M M M



NATURE | VOL 401 | 23 SEPTEMBER 1999 | www.nature.com

Confirmation that the same mechanism applies to the S49P variant of neuroserpin is provided by electron microscopy of the intraneuronal inclusions isolated from the FENIB family (Fig. 3). These show that the whole inclusions are formed by entangled fibrils which immunogold label for neuroserpin. To look more closely at the morphology of individual fibrils, the inclusions were sonically fragmented for 5 min to give short-chain filaments, and these again have the typical appearance of other serpin loop-sheet polymers



Alan Magazines Ltd

(Fig. 3). Supporting evidence for the loop-sheet basis of the linkage is provided by the biochemical properties of the fibrils, which were non-dissociable in 1 M potassium chloride but dissociated to a single band on SDS-PAGE in the absence of thiol-reducing agents. As expected for such polymers, there was a complete absence of both inhibitory activity and loop cleavage, even in the presence of a 10-fold excess of the cognate proteinase, tissue plasminogen activator. Furthermore, the histological appearance of the inclusions, as PAS-positive diastase-resistant bodies, emphasizes the identity of the general pathophysiology with the late-stage blockage of processing and secretion which results in nearly identical histology with the closely homologous variant of α_1 -antitrypsin. Finally, the clinical presentation of the two FENIB families closely matches the severity of their individual mutations. Although the domain controlling the opening of the main sheet of the molecule is centred on Ser 49 in neuroserpin, another adjacent conserved serine (56 in the template, 52 in neuroserpin) forms an even more critical interaction with the sheet¹². In keeping with the predicted gradation in severity of the molecular lesions, the neurodegeneration presents some 20 years earlier with S52R than it does in the family with the S49P mutation.

The detailed models provided from similar lesions in other serpins immediately give a clear overall molecular mechanism for FENIB that is as yet incomplete for the other familial dementias. The models from other serpins includes diseases arising from mutations in antithrombin¹⁷ and C1-inhibitor^{14,18,19} (Fig. 4), but most pertinent of all to the neuroserpin abnormality is the model

provided by the various mutations of α_1 -antitrypsin resulting in liver disease. These answer the question posed by other neurodegenerative disorders, such as the spongiform encephalopathies²⁰ and Alzheimer's disease²¹, as to whether protein deposition and accumulation is in itself sufficient to explain the late-onset dementia. The answer from α_1 -antitrypsin-associated liver disease is clear; hepatocyte loss and eventual cirrhosis is a consequence of variant deposition and not loss of function, as cirrhosis only develops with a conformationally unstable variant and not with those mutations giving complete 'null' suppression of synthesis. Where the model provided by α_1 -antitrypsin differs from that of neuroserpin is that, although in both cases a heterozygous abnormality is sufficient to produce large intracellular inclusions, it is usually only in homozygotes for α_1 -antitrypsin variants that the accumulation is severe enough to cause liver damage²². The two variants of neuroserpin, however, result in an autosomal dominant disease, consistently presenting as a dementia in young to middle-aged adults. This may reflect the grossness of the mutations with the replacement of Ser 49 by the imino acid proline or of the critical Ser 52 by arginine. The serious consequences of neuroserpin instability are relevant to the dementias as a whole because of the much greater risk that protein accumulation poses to long-lived and non-dividing cells such as neurons; the damage to neurons from such protein deposition will be cumulative and irreversible, in keeping with the observed consistency and late onset of presentation of the dementias in general. FENIB thus conforms to the general pattern of neurodegeneration caused by aberrant protein processing and tissue deposition. □

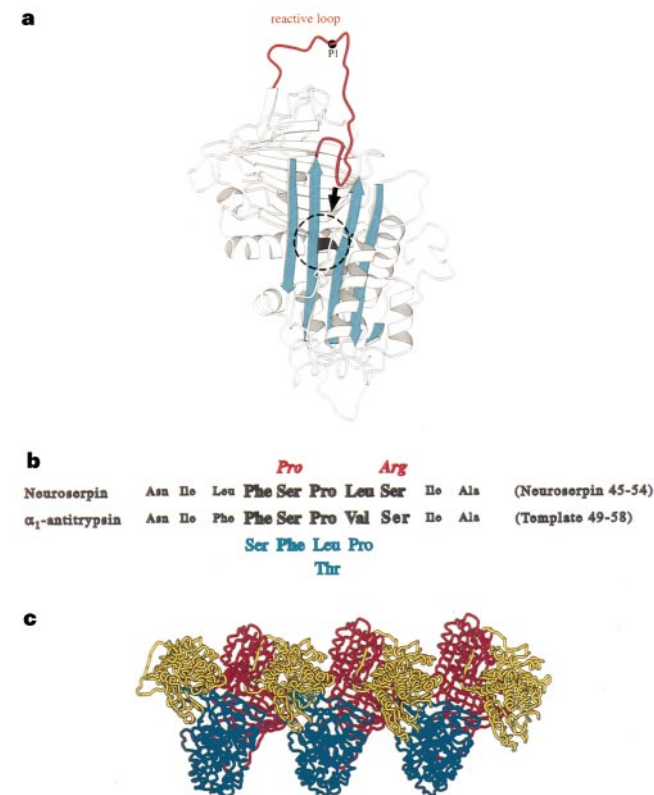


Figure 4 Serpin polymerization. **a**, Active serpin showing the initial entry of the reactive loop (red) into the partially opened A-sheet¹² (blue). The broken circle outlines the domain controlling further opening of the sheet with the critical Phe-Ser-Pro sequence (black) at the start of the B-helix. As indicated by the arrow, the loop can further insert or, in the presence of mutations, be replaced by the loop of another molecule to give a loop-sheet intermolecular linkage. **b**, The homologous sequences of the domain in neuroserpin and α_1 -antitrypsin with replacements resulting in disease (red for neuroserpins Syracuse and Portland; blue for other polymrogenic serpins). The S53P mutation in α_1 -antitrypsin Siyama is bold blue. **c**, Model of sequential loop-sheet linkage in serpin polymers (reproduced with permission²⁶).

Methods

Collins bodies

The inclusions were identified in sections of cerebral cortex by staining with haematoxylin and eosin (H and E). Brain tissue was fixed in 10% formalin, embedded in a paraffin, sectioned at 8 μ m and stained with H and E by standard procedures. The Collins bodies were isolated from 2 g of unfixed, frozen cortical tissue obtained at autopsy from an affected individual (provided by the Harvard Brain Tissue Resource Center). The isolation scheme consisted of homogenization in 10 ml 250 mM sucrose, 10 mM EDTA, 10 mM HEPES, pH 7.4, containing 10 μ l of protease inhibitor cocktail (Sigma). The bodies were collected by centrifugation at 45,000 g for 20 min, and subjected to 2 cycles of detergent washes (1% n-lauryl sarcosine), collagenase digestion (1 mg ml⁻¹), and a final wash in the homogenization medium. After each step, the bodies were collected by centrifugation and viewed microscopically after H and E staining. The bodies were solubilized by heating at 75 °C for 2 h in 4% SDS, 125 mM Tris-HCl, pH 6.8, 20% glycerol, 10% mercaptoethanol. After solubilization, the supernatant fraction was separated by SDS-PAGE in a 7.5% gel and stained for either protein with Coomassie blue or for neuroserpin using an anti-neuroserpin antibody diluted 1/2,500 in 10% blotto.

Genomic DNA

Peripheral blood was used as a source of genomic DNA, which was extracted using inorganic extraction kits (Oncor). Paraffin-embedded, formalin-fixed brain biopsy material was extracted using standard procedures.

PCR

Amplification was performed in 50–100 μ l reactions containing 500 ng genomic DNA, 0.2 mM each dNTP, 1.5 mM MgCl₂, 0.5 μ M each oligodeoxynucleotide primer, 100 μ g ml⁻¹ gelatin, 10 mM Tris-HCl, pH 8.3, 50 mM KCl. A small wax pellet was added to each PCR reaction, which was heated at 95 °C for 5 min, and cooled on ice before the addition of 10 μ l (1.5 U) *Taq* DNA polymerase (Ampli-Taq Perkin-Elmer Cetus). PCR cycling was done on a Hybaid Omnigene thermocycler using the following parameters: one cycle of 2.5 min at 94 °C, 1 min at 55 °C, and 1 min at 72 °C; 34 cycles of 1 min at 94 °C, 1 min at 55 °C (*PII2* exon 6) or 58 °C (all other *PII2* exons) and 1 min at 72 °C (with the addition of 1 s per cycle); and a single cycle of 10 min at 72 °C. Five μ l of PCR product diluted 1/3 with 95% formamide loading buffer plus dye (20 μ l total volume loaded) was heated at 85 °C for 3 min, cooled rapidly on ice, and then separated on the gel at 50 °C for between 5,000 and 15,000 Vh depending on the fragment sizes being detected²⁴. PCR-amplified products were detected by silver staining²⁴.

PCR primers from the *PII2* introns flanking the coding exons were designed from the human *PII2* genomic sequence. exon 2: (NSE2F) CAACATATCCTTCATGAGAC; (NSE2R) ACAACATAACTGAGTCAAAGTC 402-bp product; (NSE2S) (AGCTCGA-GATTC)/GCTTGAAACTGTTACAATATGGC 344-bp product. The bracketed sequence contains multiple restriction sites not present in the genomic sequence and which are not necessary if not cloning the PCR product. The use of NSE2S, rather than NSE2F, with NSE2R made the detection of the *PII2*_{Syracuse} mutation easier on 2% w/v agarose gels

following *MnII* digest. Exon 3: (NSE3F) GCTGTGCTTTAATGCTCCTCC; (NSE3R) GATACCAACTCAAATGCTCTC 435-bp product. Exon 4: (NSE4F) GGTCCCCCTTGATCTTCCAG; (NSE4R) GCCCAAATTAGATTGACAAGG 362-bp product. Exon 5: (NSE5F) TGATAGGCATCTTTTATGGCC; (NSE5R) TGATAGCC-CATCGTCCATGC 292-bp product. Exon 6: (NSE6AF) TGTTCCAGGTAACAA-GATGCTC; (NSE6AR) CAGAACTCCATAGTAAATAGAGG 245-bp product. Exon 7: (NSE7AF) CCTAGGTTTCTTTCAGTATCC; (NSE7AR) CAAATCTAGAAGAGGGGA-GAAG 242-bp product. Exon 8: (NSE8F) AGGTGGATAGGCATAGATGG; (NSE8R) AGTACTGAGGAAAAATCTCC 261-bp product. Exon 9: (NSE9F) TTAT-TATTTTGTTCACCCCTC; (NSE9R) GTTTCCAAAGTTCACCTAGAG 435-bp product. DNA sequencing of the purified PCR products was carried out by the BioResource Center, Cornell University, Ithaca, New York.

Characterization of neuroserpin isolated from patients with FENIB

SDS-PAGE, western blot analysis, the interaction of neuroserpin with tissue plasminogen activator, electron micrographs and immunogold labelling were performed as described¹⁶. Figures were produced with MOLSCRIPT²⁵.

Received 30 June; accepted 21 July 1999.

1. Lansbury, P. T. Structural neurology: are seeds at the root of neuronal degeneration? *Neuron* **19**, 1151–1154 (1997).
2. Kazizuka, A. Protein precipitation: A common etiology in neurodegenerative disorders? *Trends Genet.* **14**, 396–402 (1998).
3. Tran, P. B. & Miller, R. J. Aggregates in neurodegenerative disease: Crowds and power? *Trends Neurosci.* **22**, 194–197 (1999).
4. Carrell, R. W. & Lomas, D. A. Conformational disease. *Lancet* **350**, 134–138 (1997).
5. Carrell, R. W. & Gooptu, B. Conformational changes and disease—serpins, prions and Alzheimer's. *Curr. Opin. Struct. Biol.* **8**, 799–809 (1998).
6. Osterwalder, T., Contartese, J., Stoeckli, E. T., Kuhn, T. B. & Sonderegger, P. Neuroserpin, an axonally secreted serine protease inhibitor. *EMBO J.* **15**, 2944–2953 (1996).
7. Hastings, C. A. *et al.* Neuroserpin, a brain-associated inhibitor of tissue plasminogen activator is localized primarily in neurons. *J. Biol. Chem.* **272**, 33062–33067 (1997).
8. Seyama, K. *et al.* Siiyama (serine 53 (TCC) to phenylalanine 53 (TTC)). A new α_1 -antitrypsin-deficient variant with a mutation on a predicted conserved residue of the serpin backbone. *J. Biol. Chem.* **266**, 12627–12632 (1991).
9. Yerby, M. S., Shaw, C.-M. & Watson, J. M. D. Progressive dementia and epilepsy in a young adult: Unusual intraneuronal inclusions. *Neurology* **36**, 68–71 (1986).
10. Cervos-Navarro, & Ulrich, H. *Metabolic and Degenerative Diseases of the Central Nervous System. Pathology, Biochemistry, and Genetics* (Academic Press, San Diego, 1995).
11. Schrimpf, S. P. *et al.* Human neuroserpin (Pi12): cDNA cloning and chromosomal localization to 3q26. *Genomics* **40**, 55–62 (1997).
12. Huber, R. & Carrell, R. W. Implications of the three-dimensional structure of α_1 -antitrypsin for structure and function of serpins. *Biochemistry* **28**, 8951–8966 (1989).
13. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. The mechanism of Z α_1 -antitrypsin accumulation in the liver. *Nature* **357**, 605–607 (1992).
14. Eriksson, S., Carlsson, J. & Velez, R. Risk of cirrhosis and primary liver cancer in α_1 -antitrypsin deficiency. *N. Engl. J. Med.* **314**, 736–739 (1986).
15. Stein, P. E. & Carrell, R. W. What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Struct. Biol.* **2**, 96–113 (1995).
16. Lomas, D. A., Finch, J. T., Seyama, K., Nukiwa, T. & Carrell, R. W. Alpha₁-antitrypsin S_{iiyama} (Ser⁵³ → Phe): further evidence for intracellular loop-sheet polymerization. *J. Biol. Chem.* **268**, 15333–15335 (1993).
17. Beauchamp, N. J. *et al.* Antithrombins wobble and wobble (T85M/K): Archetypal conformational disease with *in vivo* latent-transition, thrombosis and heparin activation. *Blood* **92**, 2696–2706 (1998).
18. Aulak, K. S. *et al.* A hinge region mutation in C1-inhibitor (Ala 436 → Thr) results in nonsubstrate-like behavior and in polymerization of the molecule. *J. Biol. Chem.* **268**, 18088–18094 (1993).
19. Eldering, E., Verpy, E., Roem, D., Meo, T. & Tosi, M. COOH-terminal substitutions in the serpin C1 inhibitor that cause loop overinsertion and subsequent multimerization. *J. Biol. Chem.* **270**, 2579–2587 (1995).
20. Prusiner, S. B. Prion diseases of humans and animals. *J. R. Coll. Phys.* **28**, 1–30 (1994).
21. Yankner, B. A. New clues to Alzheimer's disease: unravelling the roles of amyloid and tau. *Nature Med.* **2**, 850–851 (1996).
22. Mahadeva, R. & Lomas, D. A. Alpha 1-antitrypsin deficiency, cirrhosis and emphysema. *Thorax* **53**, 1022–1024 (1998).
23. Shrimpton, A. E., Browitz, D. & Swender, P. CF mutation analysis in upstate New York. *Hum. Mutat.* **10**, 436–442 (1997).
24. Wallace, A. J., Williamson, P., Mountford, R. C. & Ramsden, S. C. Non-isotopic detection of microsatellite repeats using silver staining. *J. Med. Genet.* **30**, 346–347 (1993).
25. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
26. Elliott, P. R., Lomas, D. A., Carrell, R. W. & Abrahams, J. P. Inhibitory conformations of the reactive loop of α_1 -antitrypsin. *Nature Struct. Biol.* **3**, 676–681 (1996).

Acknowledgements

We thank the Harvard Brain Tissue Resource Center, J. Gusella and R. Myers, Massachusetts General Hospital for three cell lines, and C. Hubbell, C. Lamberson and K. Hicks for technical assistance. This work was supported by the Pathology Department Medical Service Group, SUNY HSC; Hendricks fund SUNY HSC (A.E.S. and P.D.H.); NIH (P.D.H.); the Wellcome Trust and Medical Research Council; and a CAP Foundation Award (R.L.D.).

Correspondence and requests for materials should be addressed to A.E.S. (e-mail: shrimpta@mailbox.hscsy.edu).

Congenital heart disease in mice deficient for the DiGeorge syndrome region

Elizabeth A. Lindsay*†, Annalisa Botta*†, Vesna Jurecic*, Sandra Carattini-Rivera*‡, Yin-Chai Cheah*, Howard M. Rosenblatt§, Allan Bradley*‡ & Antonio Baldini*†

* Department of Molecular and Human Genetics, ‡ Howard Hughes Medical Institute, and § Department of Pediatrics (Allergy and Immunology), Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

The heterozygous chromosome deletion within the band 22q11 (*del22q11*) is an important cause of congenital cardiovascular defects¹. It is the genetic basis of DiGeorge syndrome and causes the most common deletion syndrome in humans². Because the deleted region is largely conserved in the mouse, we were able to engineer a chromosome deletion (*Df1*) spanning a segment of the murine region homologous to the human deleted region. Here we describe heterozygously deleted (*Df1*+) mice with cardiovascular abnormalities of the same type as those associated with *del22q11*; we have traced the embryological origin of these abnormalities to defective development of the fourth pharyngeal arch arteries. Genetic complementation of the deletion using a chromosome duplicated for the *Df1* DNA segment corrects the heart defects, indicating that they are caused by reduced dosage of genes located within *Df1*. The *Df1*+/+ mouse model reveals the pathogenic basis of the most clinically severe aspect of DiGeorge syndrome and uncovers a new mechanism leading to aortic arch abnormalities. These mutants represent a mouse model of a human deletion syndrome generated by chromosome engineering.

Because a region of mouse chromosome 16 is highly similar in gene content and size to the *del22q11* region^{3–6} (Fig. 1a), we have been able to generate mice carrying a deletion encompassing most of this region using Cre-*loxP* chromosome engineering⁷. A proximal *loxP* site was targeted into the gene *Es2/Dgsi* (ref. 8; Fig. 1b) in mouse embryonic stem cells. The targeting vector used replaced a 200-base pair (bp) fragment containing exon 4 with the selection cassette *HprtΔ3'* (Fig. 1b; ref. 7). One of the targeted cell lines (no. 342) was used to insert the second, complementary cassette into the *Ufd1l* gene⁹, which is located about 1.2 megabases (Mb) telomeric to *Es2*. The *Ufd1l* targeting vector replaced a 4.5-kb fragment containing exons 2–3 with the *HprtΔ5'* cassette (Fig. 1c). We obtained three cell lines in which both the *Es2* and *Ufd1l* genes were replaced (nos 5, 74 and 217). Recombination between the two *loxP* sites in these cell lines was induced by transient expression of Cre, and clones that had undergone the desired recombination were selected in HAT medium as described⁷. Clones carrying a deletion on one chromosome and the complementary duplication on the other were derived from the double-targeted cell line 217, as a result of recombination between *loxP* sites located *in trans* (Figs 1d and 2a, b). We used two of these cell lines (217a and 217c) to generate chimaeras that transmitted the deletion and duplication to their progeny (Fig. 2c–e). The deletion, termed *del(16) (Es2-Ufd1l)*, is hereafter referred to as *Df1*, and the duplication, termed *dp(16) (Es2-Ufd1l)*, as *Dp1*. The double-targeted cell line 217 (before Cre recombination) was also used to generate mice with *Ufd1l* or *Es2* mutant alleles (not shown).

We examined mutant mouse lines in the mixed genetic background 129SvEvBrd (129S5) × C57BL/6. At birth, heterozygous deleted mice (*Df1*+) were recovered at the predicted mendelian ratio, but no homozygous deleted mice were recovered. In addition,

† Present addresses: Department of Pediatrics (Cardiology) Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA (E.A.L., A. Baldini); Department of Biopathology and Diagnostic Imaging, Tor Vergata University of Rome, Italy (A. Botta).

10.4% ($n = 153$) of *Df1*/ $+$ mice died at birth, compared with 3.7% of $+/+$ littermates ($n = 133$, $P < 0.05$). Deleted mice surviving the first day of life were viable and fertile and grew normally.

To understand whether heart defects may explain the neonatal mortality, we established matings between *Df1*/ $+$ males and $+/+$ females and collected embryos at 18.5 days post coitum (d.p.c.). We examined 11 litters (77 mice), 42 (54.5%) of which were *Df1*/ $+$ and 35 (45.5%) of which were $+/+$. Eleven *Df1*/ $+$ embryos (26%) had cardiovascular abnormalities (Table 1). No abnormalities were found in $+/+$ littermates. Three embryos had interrupted aortic arch type B (IAA-B), one of which also had a ventricular septal defect (VSD) (Fig. 3b, g). One embryo had infundibular pulmonary stenosis and a VSD, and one had aberrant origin of the right subclavian artery (ARS) from the pulmonary trunk (Fig. 3c, d). The most frequent abnormality (found in six individuals) was a retro-oesophageal right subclavian artery, which originated from the descending aorta, dorsal to the emergence of the left subclavian artery (Fig. 3e). This type of ARS was found as a single defect in four cases, but was associated with overriding aorta and VSD in one case (Fig. 3h) and with an abnormal vessel connecting the pulmonary trunk with the left subclavian artery in another (not shown). We also examined adult *Df1*/ $+$ animals and found that 18% ($n = 56$) had cardiovascular abnormalities. The defect was retro-oesophageal ARS in all cases except one, which was a case of right aortic arch (RAA). Overall, *Df1*/ $+$ mice present with the cardiovascular abnormalities most frequently found in *del22q11* patients, which are: tetralogy of Fallot (a complex comprising overriding aorta, ventricular septal defects, infundibular pulmonary stenosis and acquired right ventricular hypertrophy) present in 17% of deleted patients,

IAA-B (14%) and VSD (14%)¹⁰. Although we did not find tetralogy of Fallot in *Df1*/ $+$ mice, its structural components were present.

IAA-B and ARS are embryologically related, in that the fourth pharyngeal arch arteries give rise on the right to the proximal portion of the right subclavian artery, and on the left to the aortic arch segment that is missing in IAA-B. RAA can be explained by failure of the left fourth arch artery, which is then compensated for by enlargement of the right fourth arch artery and abnormal persistence of the right dorsal aorta. To investigate whether *Df1*/ $+$ animals present with fourth pharyngeal arch abnormalities, we examined 11.5-d.p.c. embryos after intracardiac injection of India ink. Arch arteries 3, 4 and 6 should be completely formed at this stage (Fig. 4a). In five of the eleven *Df1*/ $+$ embryos examined (45%), the fourth arch arteries were either undetectable (Fig. 4b) or reduced in size (Fig. 4c, d). In most embryos, only the right arch was affected, but in some the defect was bilateral. The third and sixth arch arteries appeared normal. Wild-type littermates were normal ($n = 16$).

As DGS patients may have hypocalcaemia, T-cell immune defects and craniofacial abnormalities¹⁰, we analysed these features in our mutants. We found normal serum levels of calcium and phosphorus ($n = 13$ *Df1*/ $+$ and 9 $+/+$) and parathyroid hormone ($n = 16$ *Df1*/ $+$ and 4 $+/+$) in 2–6-month-old mice. The percentage of B and T cells in peripheral blood and splenic mononuclear cells, and lymphocyte proliferative response to mitogens, were normal in 2–6-month-old mice ($n = 16$ *Df1*/ $+$ and 9 $+/+$). The thymus was normal in size and morphology in 18.5-d.p.c. embryos and in adult *Df1*/ $+$ mice ($n = 114$ in total). Skeletal preparations were examined in three newborn litters ($n = 7$ *Df1*/ $+$ and 13 $+/+$ mice) and no gross

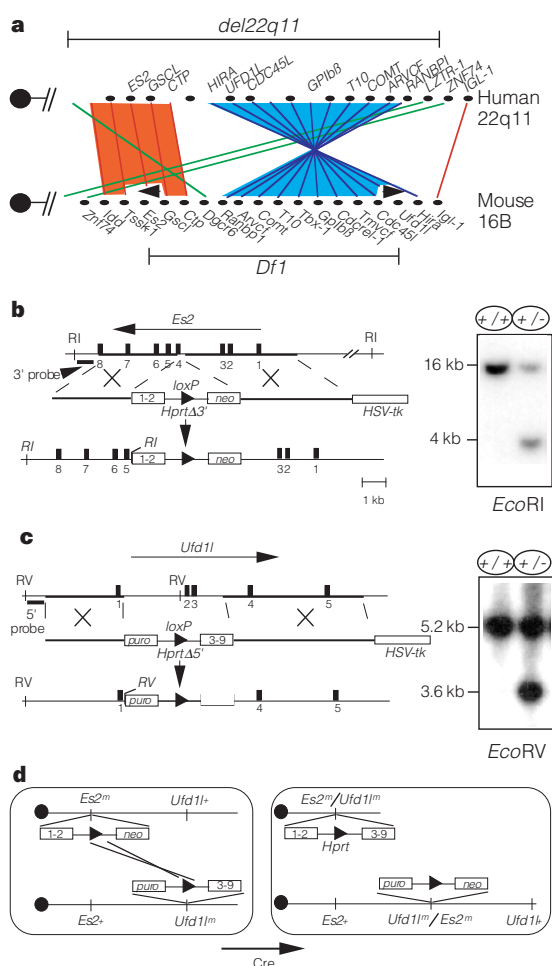


Figure 1 Comparative mapping and targeting of the murine region corresponding to the *del22q11* region. **a**, Gene content and order in the region commonly deleted in *del22q11* patients, compared with that of a region of mouse chromosome 16. Red lines indicate genes with the same order and orientation in the two species, green lines indicate genes that have changed position during evolution and blue lines indicate genes that have conserved their relative order but with inverted centromere–telomere orientation. *CLTCL* is not present in mouse. Arrowheads indicate the transcriptional orientation of *Es2* and *Ufd1l*, the two anchor points of the *Df1* targeted deletion. This scheme has been compiled on the basis of published data^{3–6} and our unpublished results. Not to scale. **b**, Top, the wild-type *Es2* chromosomal locus. 8 out of 10 exons are indicated by small rectangles. The targeting vector is shown below the wild-type allele. The chromosome engineering cassette *HprtΔ3'* contains, from left to right, a PGK promoter, the first 2 exons of a human *Hprt* minigene²⁰, an intron with a *loxP* site, and a *Poll*–*neo* cassette⁷. Below this is the targeted allele (*Es2^m*) after homologous recombination. On the right is the diagnostic restriction pattern of a recombined cell line DNA digested with *EcoRI* and probed with the 3' probe. *R*: *EcoRI* restriction sites. **c**, Wild-type, targeting vector and targeted allele (*Ufd1l^m*) of the *Ufd1l* locus. Only 5 of the 11 exons are indicated. The chromosome engineering cassette *HprtΔ5'* (ref. 7) contains a PGK–*puro* cassette (in the opposite orientation to the *Ufd1l* gene), an *Hprt* minigene intron containing a *loxP* site and exons 3–9 of the human *Hprt* minigene. Exons 2 and 3 are the first two coding exons of the *Ufd1l* gene. On the right is the diagnostic restriction pattern of a recombined cell line DNA digested with *EcoRV* and probed with the 5' probe. *R*: *EcoRV* restriction sites. **d**, Scheme showing the products of Cre-induced *trans* recombination between the *loxP* sites contained in the two complementary cassettes. ES cells with recombined chromosomes (right) are selected for *Hprt* function using HAT. These cells are also resistant to puromycin and G418. AB2.2 ES cells are *Hprt* deficient.

Table 1 Cardiovascular defects observed in *Df1*/+ 18.5 dpc embryos.

Defect	Number of embryos*
IAA-B	2
IAA-B + VSD + cervical RSA	1
ARS (DAo)	4
ARS (DAo) + extra vessel†	1
ARS (DAo) + VSD + overriding aorta	1
ARS (PT)	1
Infundibular pulmonary stenosis + VSD	1

* The total number of *Df1*/+ embryos examined is 42. No abnormalities were found in the 33 wild-type littermates examined.

† See text for details.

IAA-B: Interrupted aortic arch type B. VSD, ventricular septal defect. RSA, right subclavian artery. ARS (DAo), Aberrant right subclavian artery originating from descending aorta. ARS (PT), Aberrant subclavian artery originating from pulmonary trunk.

abnormality was found. No cleft palate or other gross palatal abnormalities were found in 18.5-d.p.c. *Df1*/+ embryos ($n = 42$).

To test whether cardiovascular abnormalities in *Df1*/+ mice are due to haploinsufficiency of deleted genes or to an effect of the deletion on neighbouring genes, we genetically complemented the deletion using the *Dp1* duplication. This duplication will compensate for genes deleted in *Df1* but will not compensate for an effect of the deletion on neighbouring genes, or for the deletion of anchor-point genes (*Es2* and *Ufd1l*; Fig. 1d). We examined 55 embryos (18.5 d.p.c.) from matings between *Df1*/+ males and *Dp1*/+ females. Cardiovascular abnormalities were found in 5 (30%) of the 16 *Df1*/+ embryos, as expected, but in none of the 15 *Df1*/*Dp1* embryos. An additional 24 *Df1*/*Dp1* embryos were obtained from matings between *Df1*/+ males and *Dp1*/*Dp1* females, and again no heart

defects were found. Also, no neonatal mortality was observed among *Df1*/*Dp1* mice ($n = 43$) and no heart defects were found in 10 adult *Df1*/*Dp1* mice examined. Hence, the *Df1*/+ phenotype is due to haploinsufficiency of genes located within the deletion. Because *Df1*/*Dp1* mice have only one functional allele of genes *Es2* and *Ufd1l*, we can also conclude that heterozygous mutation of these genes is insufficient to generate heart defects. The role of the *Es2* and *Ufd1l* genes in mammalian development is unknown. However, no homozygous *Es2* or *Ufd1l* mutants were recovered at 9.5 d.p.c., indicating that these genes are required for early mouse embryonic development (data not shown). *Ufd1l* has been proposed as a candidate gene for the *del22q11* phenotype^{9,11}. If heterozygous deletion of *Ufd1l* were the sole cause of heart defects in *Df1*/+ mice, we would expect *Ufd1l*^{+/-} mice to have the same heart abnormalities. This is in contrast to the negative results obtained with *Df1*/*Dp1* mice. In addition, we examined 30 *Ufd1l*^{+/-}, 18.5-d.p.c. embryos and found no heart abnormalities. Neonatal mortality among *Ufd1l*^{+/-} mice ($n = 117$) was no different from that of wild-type littermates. The lack of heart defects in both *Ufd1l*^{+/-} and *Df1*/*Dp1* mice is unlikely to be due to residual function of the *Ufd1l* mutant alleles because, in *Ufd1l*^{+/-} embryos, the *Ufd1l* message is reduced to about 50% of the +/+ level (Fig. 2f). Furthermore, in *Df1*/*Dp1* animals the *Ufd1l* mutant allele is different because the gene is not

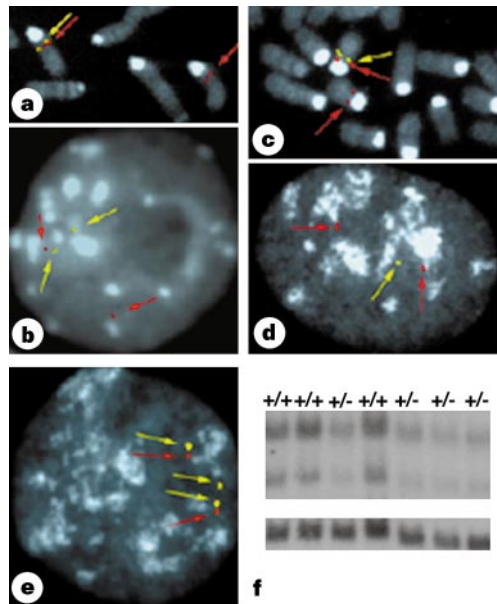


Figure 2 *Df1* deletion and *Dp1* duplication in ES cells and mice, and *Ufd1l* expression in *Ufd1l*^{+/-} mice. Examples of fluorescence *in situ* hybridization experiments on chromosomes from a deleted/duplicated ES cell line (a, b) and from primary explant fibroblasts of a mouse carrying the chromosome deletion *Df1* (c, d) and one carrying the duplication *Dp1* (e). Control probes are shown in yellow and the test probe is shown in red. Two control probes were utilized; one for the *Hira* locus that is distal to *Df1* (a, b, e) and one for the *scid* locus that is proximal to *Df1* (c, d). The test probe identifies the *Comt* locus within the deleted or duplicated segment. f, Northern blotting analysis of total RNA (~10 µg per lane) from *Ufd1l*^{+/+} and *Ufd1l*^{+/-} 10.5 d.p.c. embryos. The filter was hybridized with a *Ufd1l* complementary DNA (top) and, after signal removal, with a *Gapdh* probe (control, bottom). The *Ufd1l* probe detects two bands deriving from alternative usage of polyadenylation signals. The intensity of both *Ufd1l* bands is reduced in *Ufd1l*^{+/-} embryo RNA. Densitometric analysis of the bands, after equalization with the control signal, indicates a ~50% reduction in intensity of the signal in *Ufd1l*^{+/-} embryos.

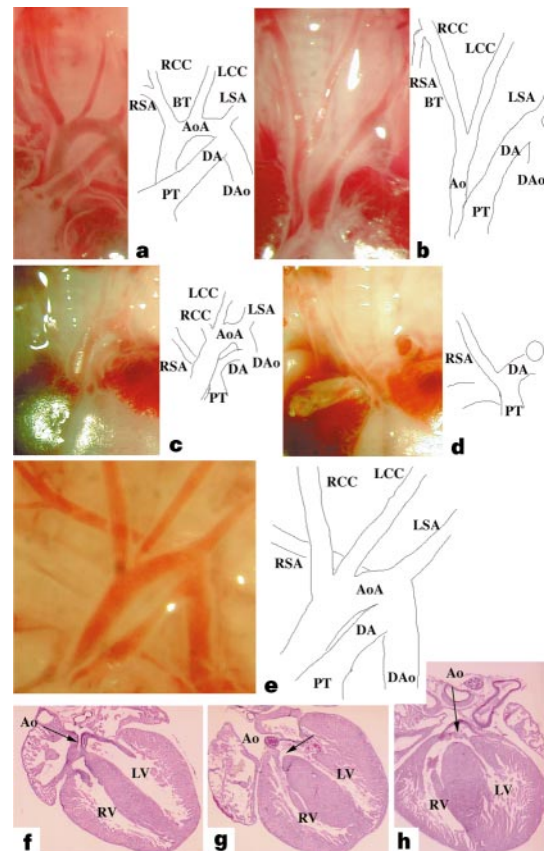


Figure 3 Examples of cardiovascular findings in 18.5 d.p.c. *Df1*/+ embryos. a, Normal anatomy in a wild-type embryo. b, IAA-B with distal (cervical) emergence of the RSA. c, Aberrant origin of the RSA; d, the same individual after removal of the aortic arch to show the origin of the RSA from the pulmonary trunk. e, Aberrant origin of the RSA from the descending aorta. f, Histological section of a wild-type 18.5 d.p.c. embryo. g, Heart from the individual shown in b; arrow indicates a ventricular septal defect in the membranous portion of the septum. h, Aberrant position of the emergence of the aorta from above the right and left ventricle (overriding aorta) in another *Df1*/+ animal. Ao, aorta; AoA, aortic arch; DAo, descending aorta; BT, brachiocephalic trunk; DA, ductus arteriosus; PT, pulmonary trunk; RSA and LSA, right and left subclavian arteries; RCC and LCC, right and left common carotid arteries.

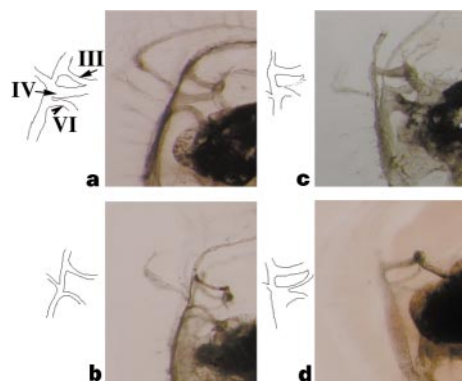


Figure 4 Right aortic arch arteries 3, 4 and 6 visualized by intracardiac injection of India ink in 11.5 d.p.c. embryos. **a**, Normal anatomy; the third, fourth and sixth arches are indicated as III, IV and VI, respectively. Note the relatively larger size of the fourth arch artery. **b–d**, In these *Df1*^{+/+} embryos, the right fourth arch artery is not detectable (**b**), barely visible (**c**) or reduced in size relative to the third arch artery (**d**).

only partially deleted (as in *Ufd1l*^{+/−} mutants), but is also truncated as the result of Cre recombination (Fig. 1d).

The DGS candidate gene *Hira*, encoding a transcription factor interacting with *Pax3* (which, in turn, is expressed in cardiac neural crest cells^{12,13}), lies outside *Df1*. *Hira* antisense oligonucleotides induce persistent truncus arteriosus (PTA) in chick embryos¹⁴, a defect that is observed in 9% of *del22q11* patients¹⁰. PTA was not found in *Df1*^{+/+} mice, which indicates that there may be at least two heart-development-related loci within *del22q11*.

Aortic arch patterning defects have been shown to derive from inadequate formation or remodelling of pharyngeal arch arteries 3, 4 and 6, but there are considerable differences between models. For example, in the chick, ablation of cardiac neural crest affects all three arch arteries, which form normally but regress rapidly¹⁵. In mice lacking the *Endothelin-1*, *Endothelin Converting Enzyme-1* or *Endothelin-A* receptor genes, the fourth and sixth arch arteries regress to a variable extent beginning at 11.5 d.p.c.^{16,17}. In mice lacking the *MFH-1* gene, the fourth arch arteries are specifically affected, but not before 12.5 d.p.c., when they fail to remodel¹⁸. In contrast, in *Df1*^{+/+} mice, the fourth arch arteries are affected specifically and the abnormalities are already present at 11.5 d.p.c., that is, before the asymmetric remodelling of these arteries occurs. These results indicate that the defects in *Df1*^{+/+} mice may be due to inadequate formation, early regression or growth failure of the fourth aortic arch arteries. We can speculate that *Df1* affects a local signalling pathway responsible for growth, survival and/or differentiation of neural crest- or endoderm-derived cells of the fourth arch. Because the embryonic lesion is restricted to the fourth arch, a primary defect of cardiac neural crest cell migration is unlikely. This would explain why one of the most typical heart defects resulting from cardiac neural crest ablation, PTA, and phenotypic findings that are presumed to be of neural crest origin (such as hypoparathyroidism, thymic and craniofacial abnormalities), are not found in *Df1*^{+/+} mice. The genes responsible for these abnormalities may lie outside *Df1*, or heterozygous mutations may be insufficient to generate these defects in mice, or in the genetic background tested.

Although not covering the entire phenotypic spectrum of DiGeorge syndrome, *Df1*^{+/+} mice are a model of the most marked clinical finding in this syndrome, and they provide insight into the embryological mechanisms underlying congenital heart disease in DiGeorge syndrome patients. These mutants illustrate an approach that can be used to genetically dissect deletion syndromes¹⁹ and represent, to our knowledge, the first animal model of this relatively

common type of genetic disease obtained using chromosome engineering. □

Methods

Targeting experiments and mouse breeding

We constructed *Es2* and *Ufd1l* replacement targeting vectors using the previously described chromosome engineering cassettes *HprtΔ3'* and *HprtΔ5'*, respectively⁷. The two cassettes include a selectable marker gene (*neo* and *puro*, respectively), a *loxP* site and exons 1–2 and 3–9, respectively, of a human *Hprt* minigene. Vector DNA was electroporated into AB2.2 embryonic stem (ES) cells derived from *Hprt*-deficient 129SvEvBrd²⁰. Cells electroporated with the *Es2* vector were selected with G418 and FIAU (1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil) for 7–9 days. We isolated 288 G418- and FIAU-resistant clones and screened them by Southern blotting using the probe *Es2*–3'. About 50% of these clones showed the diagnostic *EcoRI* restriction pattern (Fig. 1b). Cells from one of these clones were electroporated with *Ufd1l* vector DNA. We isolated 288 puromycin- and FIAU-resistant clones and analysed them by Southern blotting. Three clones, (nos 5, 74 and 217) were recombinant and carried both *Es2* and *Ufd1l* mutated alleles. For each of these clones, 10⁷ cells were electroporated with the Cre expression plasmid pOG231 to catalyse recombination between *loxP* sites. The chromosome engineering cassettes are designed to produce a functional *Hprt* gene after *loxP* recombination. Electroporated cells were selected with HAT (10 mM sodium hypoxanthine, 40 μM aminopterin, 1.6 mM thymidine), which kills *Hprt*-deficient cells, hence providing a positive selection for the *loxP* recombination event. Clones 5 and 217 produced a recombination frequency of 1:30,000 electroporated cells, whereas for clone 74 the frequency was 1:50. HAT-resistant clones from 5 and 217 cells were also resistant to G418- and puromycin (recombination *in trans*, Fig. 1d) whereas clones from 74 were G418- and puromycin-sensitive (recombination *in cis*). Cell line 217 (double-targeted *in trans*) and two recombined cell lines, 217a and 217c (deleted/duplicated) were injected into C57BL/6 blastocysts to produce chimaeric mice. Male chimaeras were crossed with C57BL/6 females and all the mutated alleles transmitted to the progeny. Fluorescence *in situ* hybridization of chromosomes from cell lines and primary fibroblasts of mouse tail biopsies was performed as described²¹. Southern and northern blotting analyses were performed according to standard procedures.

Procedures used for phenotypic analyses

For histological examination, samples were fixed in PBS-buffered 4% formaldehyde, embedded in paraffin, cut in 7-μm-thick sections and stained with haematoxylin and eosin, according to standard procedures²². For India ink injection, embryos were collected and injected intracardially with ink, and fixed in PBS-buffered 4% formaldehyde. Fixed embryos were dehydrated through a graded series of ethanol and cleared in 1:1 benzyl benzoate: methyl salicylate. Craniofacial and skeletal morphology was examined by gross inspection and by analysis of the skeleton of newborns after staining with alizarin red S and alcian blue, and clearing with KOH as described²³. We measured serum levels of parathyroid hormone using a commercial immunoradiometric assay (Nichols Institute Diagnostics). Lymphocyte phenotype was analysed using a whole-blood lysis method, monoclonal antibody staining and standard two-colour flow cytometry using a Coulter EPICS II flow cytometer²⁴. The surface markers tested were CD3, CD4 and CD8 (for T cells) and CD19 for B cells. Lymphocyte proliferation response to phytohaemagglutinin and concanavalin-A was assayed using [³H]thymidine incorporation in 3- and 4-day cultures, using standard procedures^{25,26}.

Received 26 March; accepted 19 July 1999.

- Lindsay, E. A. & Baldini, A. Congenital heart defects and 22q11 deletions: which genes count? *Mol. Med. Today* **4**, 350–358 (1998).
- Wilson, D. I. *et al.* Minimum prevalence of chromosome 22q11 deletions. *Am. J. Hum. Genet.* (Suppl.) **55**, A975 (1994).
- Galili, N. *et al.* A region of mouse chromosome 16 is syntenic to the DiGeorge, velocardiofacial syndrome minimal critical region. *Genome Res.* **7**, 17–26 (1997).
- Botta, A., Lindsay, E. A., Jurecic, V. & Baldini, A. Comparative mapping of the DiGeorge Syndrome region in mouse shows inconsistent gene order and differential degree of gene conservation. *Mamm. Genome* **8**, 890–895 (1997).
- Puech, A. *et al.* Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization. *Proc. Natl. Acad. Sci. USA* **94**, 14608–14613 (1997).
- Sutherland, H. F., Kim, U. J. & Scambler, P. J. Cloning and comparative mapping of the DiGeorge syndrome critical region in the mouse. *Genomics* **52**, 37–43 (1998).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Lindsay, E. A., Harvey, E. L., Scambler, P. J. & Baldini, A. *ES2*, a gene deleted in DiGeorge syndrome, encodes a nuclear protein and is expressed during early mouse development, where it shares an expression domain with a Goosecoid-like gene. *Hum. Mol. Genet.* **7**, 629–635 (1998).
- Pizzuti, A. *et al.* UFD1L, a developmentally expressed ubiquitination gene, is deleted in CATCH 22 syndrome. *Hum. Mol. Genet.* **6**, 259–265 (1997).
- Ryan, A. K. *et al.* Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *J. Med. Genet.* **34**, 798–804 (1997).
- Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T. & Srivastava, D. A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science* **283**, 1158–1161 (1999).
- Magnaghi, P., Roberts, C., Lorain, S., Lipinski, M. & Scambler, P. J. *HIRA*, a mammalian homologue of *Saccharomyces cerevisiae* transcriptional co-repressors, interacts with *Pax3*. *Nature Genet.* **20**, 74–77 (1998).
- Conway, S. J., Henderson, D. J. & Copp, A. J. *Pax3* is required for cardiac neural crest migration in the mouse: evidence from the splotch (*Sp*^{2H}) mutant. *Development* **124**, 505–514 (1997).

14. Farrell, M. J. *et al.* HIRA, a DiGeorge syndrome candidate gene, is required for cardiac outflow tract septation. *Circ. Res.* **84**, 127–135 (1999).
15. Kirby, M. L. in *Heart Development* (eds Harvey, R. P. & Rosenthal, N.) 179–193 (Academic, San Diego, 1999).
16. Kurihara, Y. H. *et al.* Aortic arch malformations and ventricular septal defects in mice deficient in Endothelin-1. *J. Clin. Invest.* **96**, 293–300 (1995).
17. Yanagisawa, H. *et al.* Role of Endothelin-1/Endothelin-A receptor-mediated signaling pathway in the aortic arch patterning in mice. *J. Clin. Invest.* **102**, 22–33 (1998).
18. Iida, K. *et al.* Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis. *Development* **124**, 4627–4638 (1997).
19. Liu, P., Zhang, H., McLellan, A. Vogel & Bradley, A. Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* **150**, 1155–1168 (1998).
20. Matzuk, M. M., Finegold, M. J., Su, J. G., Hseuh, A. J. & Bradley, A. α -inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature* **360**, 313–319 (1992).
21. Baldini, A. & Lindsay, E. A. in *In Situ Hybridization Protocols* (ed. Choo, K. H. A.) 75–85 (Humana, Clifton, New Jersey, 1994).
22. Kaufman, M. H. *Atlas of Mouse Development* (Academic, San Diego, 1992).
23. McLeod, M. J. Differential staining of cartilage and bone in whole mouse fetuses by alcian blue and alizarin red S. *Teratology* **22**, 299–301 (1980).
24. Kidd, P. M. & Nicholson, J. K. in *Manual of Clinical Laboratory Immunology* 5th edn (eds Rose, N. R., deMacario, E. C., Folds, J. D., Lane, H. C. & Nakamura, R. M.) 229–244 (ASM, Washington DC, 1997).
25. Kruisbeek, A. M. & Shevach, E. in *Current Protocols in Immunology* Sections 3.12.1–3.12.14 (John Wiley and Sons, New York, 1991).
26. Kruisbeek, A. M. in *Current Protocols in Immunology* Sections 3.1.1–3.1.5 (John Wiley and Sons, New York, 1993).

Acknowledgements

We thank D. Su, Y. Wang and N. R. Parikh for technical assistance. This work was funded by grants from the NIH. A. Bradley is an investigator with the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to A. Baldini (e-mail: baldini@bcm.tmc.edu).

Polycystin-L is a calcium-regulated cation channel permeable to calcium ions

Xing-Zhen Chen^{*†}, Peter M. Vassilev^{†‡}, Nuria Basora^{*}, Ji-Bin Peng^{*}, Hideki Nomura^{*}, Yoav Segal^{*}, Edward M. Brown[‡], Stephen T. Reeders^{*}, Matthias A. Hediger^{*} & Jing Zhou^{*}

^{*} Renal and [‡] Endocrine-Hypertension Divisions and Membrane Biology Program, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

[†] These authors contributed equally to this work

Polycystic kidney diseases are genetic disorders in which the renal parenchyma is progressively replaced by fluid-filled cysts¹. Two members of the polycystin family (polycystin-1 and -2) are mutated in autosomal dominant polycystic kidney disease (ADPKD)^{2–5}, and polycystin-L is deleted in mice with renal and retinal defects⁶. Polycystins are membrane proteins that share significant sequence homology^{6,7}, especially polycystin-2 and -L (50% identity and 71% similarity). The functions of the polycystins remain unknown. Here we show that polycystin-L is a calcium-modulated nonselective cation channel that is permeable to sodium, potassium and calcium ions. Patch-clamp experiments revealed single-channel activity with a unitary conductance of 137 pS. Channel activity was substantially increased when either the extracellular or intracellular calcium-ion concentration was raised, indicating that polycystin-L may act as a transducer of calcium-mediated signalling *in vivo*. Its large single-channel conductance and regulation by calcium ions distinguish it from other structurally related cation channels.

Polycystin-L (PCL) and polycystin-2 share similarities in sequence, domain organization and/or membrane topology⁶

(Fig. 1a) with cation channels such as the transient receptor potential (TRP) and voltage-gated Ca^{2+} , Na^{+} and K^{+} channel families^{5,6,8–10}. Therefore, we utilized the *Xenopus* oocyte expression system to test whether PCL functions as a channel.

Oocytes injected with RNA encoding full-length PCL exhibited a 10-fold increase in total conductance, compared to water-injected oocytes, in NaCl-containing solution ($18.5 \pm 0.8 \mu\text{S}$ versus $1.9 \pm 0.2 \mu\text{S}$; Fig. 1b). Identical outward currents were obtained with 50 or 100 mM external Cl^{-} (Fig. 1c), suggesting that they are due to cation (predominantly K^{+}) efflux and not to Cl^{-} influx. Inward currents were largely abolished by replacing external Na^{+} with *N*-methyl-D-glucamine (NMDG; Fig. 1d). PCL exhibited similar permeabilities (P) to Na^{+} , K^{+} and Rb^{+} , lower permeability to Li^{+} ($P_{\text{Na}} : P_{\text{K}} : P_{\text{Rb}} : P_{\text{Li}} = 1 : 0.98 : 0.97 : 0.87$), and very low permeabilities to large cations such as NMDG, tetraethylammonium or choline (Fig. 1d).

Single-channel analysis of 102/127 cell-attached patches of PCL-injected oocytes, but not 50/50 H_2O -injected oocytes or those

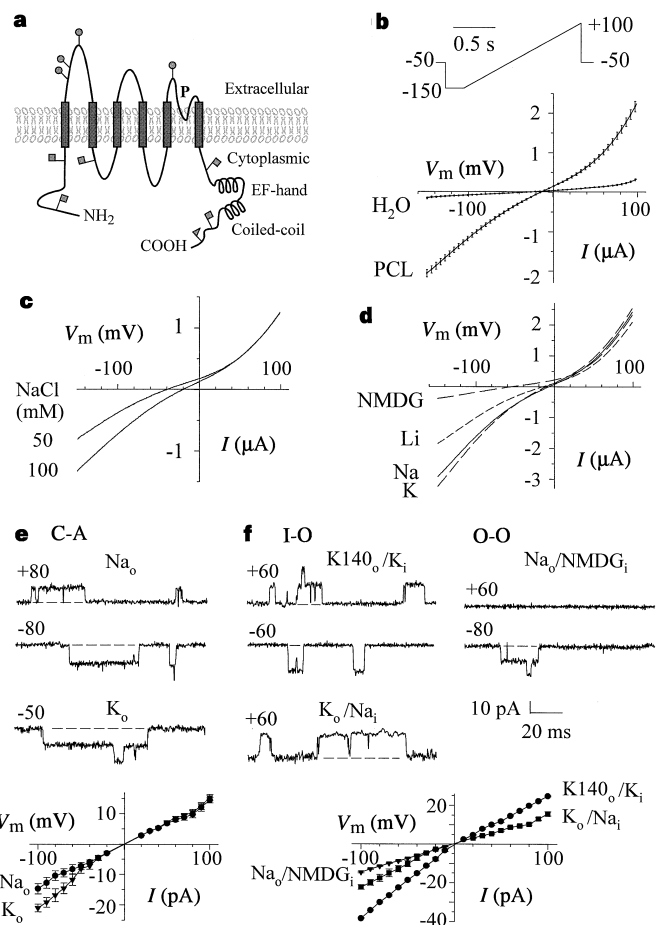


Figure 1 PCL channel conductance in the resting state. **a**, Proposed membrane topology for PCL. 'P' and circular, triangular and square flags indicate, respectively, the pore region, N-glycosylation, PKA- and PKC-phosphorylation sites. **b**, Average current–voltage (*I*–*V*) curves obtained from H_2O -injected ($n = 16$) or PCL-expressing ($n = 47$) oocytes, following a voltage ramp (top) in NaCl-containing solution. **c**, Comparison of currents obtained from the same oocyte in NaCl-containing solution and when 100 mM mannitol replaced 50 mM NaCl. **d**, Ion-selectivity *I*–*V* curves obtained using cation-Cl-substituted solutions. 2 mM KCl was removed from all solutions. **e**, Upper panels, selected recordings of PCL single-channel currents at indicated voltages (in mV) in the cell-attached (C-A) configuration. Dashed lines represent currents of the close state. Lower panel, average *I*–*V* relations ($n = 5$). The bath solution was 'Na'. **f**, Upper panels, PCL channel activities in inside-out (I-O) or outside-out (O-O) patches. Lower panel, average *I*–*V* relations ($n = 5$ –10) obtained in I-O and O-O modes.

expressing other membrane proteins ($n = 40$), revealed a channel that is highly permeable to Na^+ and K^+ (Fig. 1e). The main unitary conductance at ± 50 mV in the presence of 100 mM external Na^+ was 137 ± 7 pS ($n = 8$). In the presence of external NaCl, excised patches only exhibited inward currents when intracellular cations were replaced by NMDG (Fig. 1f), indicating that PCL is impermeable to Cl^- and NMDG. Our data indicate that the outward currents observed in the cell-attached and whole-cell configurations were carried mainly by K^+ , and that PCL functions as a nonselective cation channel.

We investigated divalent cation permeation, and found a multiphasic response when we raised extracellular Ca^{2+} (Ca_o^{2+}). In PCL-expressing oocytes pre-incubated in Ca^{2+} -free solutions for 3–5 min, large currents were evoked by adding 5 mM Ca_o^{2+} (Fig. 2a, b). At -50 mV, currents reached a peak 15–30 s after Ca_o^{2+} was added and then decreased to below basal levels with a time constant of 0.6 ± 0.1 min ($n = 17$; Fig. 2a). Ca^{2+} entry, determined by ^{45}Ca uptake, was markedly higher during the first 3-min period than during the second period (Fig. 2b). However, re-applying 5 mM Ca_o^{2+} inhibited inward currents. Thus, applying Ca_o^{2+} evoked activation and subsequent desensitization of PCL conductance, as assessed by whole-cell current and Ca^{2+} entry into oocytes. When Ca_o^{2+} was removed immediately after activation, the conductance during desensitization (dashed curve) was larger than before activation (point 1, Fig. 2a). Desensitization was, however, reversible, as Ca^{2+} -induced activation recovered after 5–10 min incubation without Ca_o^{2+} .

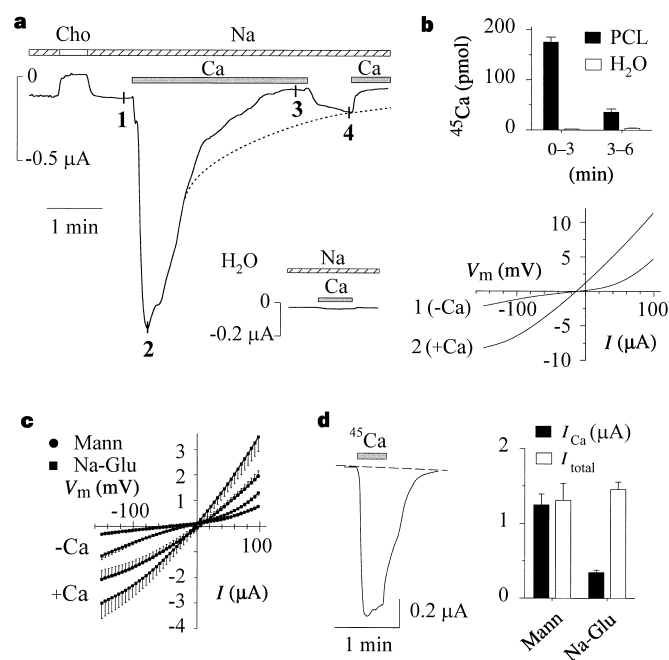


Figure 2 Polycystin-L conductance is stimulated by Ca^{2+} . **a**, Whole-cell currents recorded at -50 mV in the presence of choline-Cl-substituted (Cho) and NaCl-containing (Na) solutions. Addition of 5 mM CaCl_2 is indicated by grey bars. The dotted curve approximates the time course of the inward current when external Ca^{2+} was removed after the current reached its peak (point 2). The inset is a recording at -50 mV from an H_2O -injected oocyte. **b**, Top, ^{45}Ca uptake in oocytes pre-incubated in NaCl-containing solution for 10 min. Non-radioactive Ca^{2+} (5 mM) was added at time 0. 0–3 and 3–6 indicate the periods (in min) of ^{45}Ca incubation. Bottom, $I-V$ curves obtained from the same oocyte before (point 1) or after (point 2) Ca^{2+} addition. **c**, Average $I-V$ curves obtained before or after external Ca^{2+} addition, in Na-glutamate- (Na-Glu) or mannitol (Mann)-substituted solution, in the same oocytes ($n = 10$) pre-incubated in Cl^- -free solution for >10 h. **d**, Left, current at -50 mV in mannitol-substituted solution before and after activation by 5 mM $^{45}\text{CaCl}_2$. Ca^{2+} uptake was then converted into current by taking the half-width of the spike as the uptake time. Right, calcium and total currents converted similarly in mannitol- or Na-glutamate-substituted solution ($n = 4$).

tion without Ca_o^{2+} .

To determine whether Ca^{2+} induces PCL-specific activation, we eliminated endogenous Ca^{2+} -activated Cl^- currents^{11,12} by pre-incubating oocytes in Cl^- -free solutions (Cl^- replaced by Hepes, gluconate or glutamate) for 10–24 h and using Na-glutamate- or mannitol-substituted solutions during measurements. Substantial increases in Ca^{2+} -activated current still occurred (Fig. 2c), indicating that the PCL-specific cation conductance was increased by Ca_o^{2+} . Hereafter, we define PCL channel activity as being in the resting, activated or desensitized state under the conditions corresponding to points 1, 2 and 3 in Fig. 2a, respectively.

As PCL is nonselectively cation permeable, we estimated the contributions of Ca^{2+} and Na^+ to the observed Ca^{2+} -activated inward current by simultaneously measuring whole-cell current and ^{45}Ca influx under voltage clamp (Fig. 2d, left panel). In oocytes deprived of Cl^- by incubation in Cl^- -free solutions, Ca^{2+} entry accounted for most of the total Ca^{2+} -activated current in mannitol-substituted solution and only 20% in Na-glutamate-substituted solution (Fig. 2d, right panel), indicating that Na^+ and Ca^{2+} may compete for a common permeation pathway. In the latter case, the remaining 80% of Ca^{2+} -activated currents must result from Na^+ entry. Thus, with 100 mM Na^+ and 5 mM Ca^{2+} in the external solution, PCL was about five times more permeable to Ca^{2+} than to Na^+ , at -50 mV in the activated state. In the resting state, with NaCl-containing solution, PCL was 4.3 ± 0.7 ($n = 10$) times more permeable to Ca^{2+} than to Na^+ .

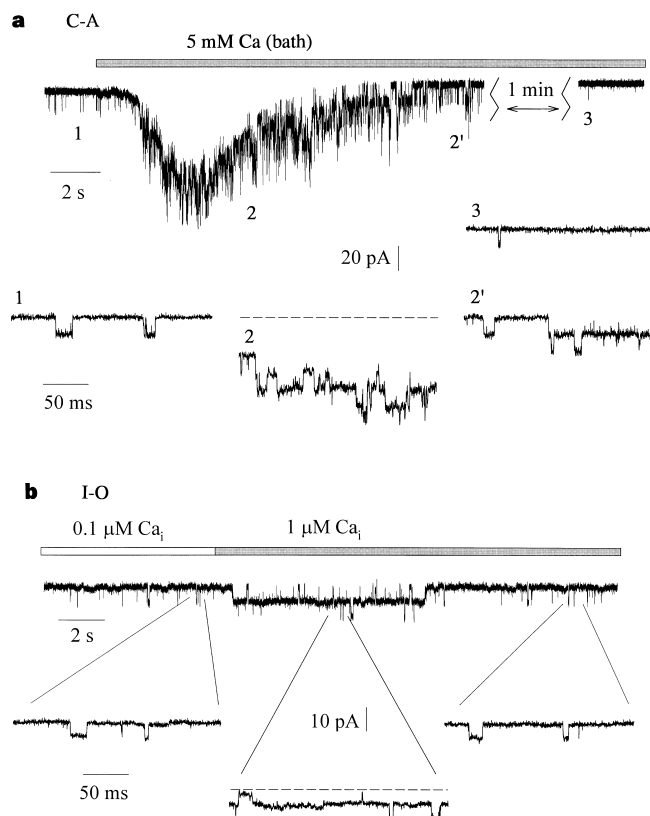


Figure 3 Activity of PCL single channels is increased by Ca^{2+} . **a**, Top, single-channel currents recorded at -60 mV in a cell-attached patch before and after adding 5 mM Ca^{2+} to the bath (Na_o solution). The pipette solution was K140. Bottom, expanded traces representing, respectively, the resting (1), activated (2, 2') and desensitized (3) states. Resting potential of oocytes was assumed to be -30 mV. **b**, Representative traces at -50 mV in an inside-out patch before and after adding 1 μM Ca_i^{2+} (top). Selected traces are expanded (bottom). Open probabilities were 0.14 ± 0.03 , 0.58 ± 0.12 , and 0.10 ± 0.02 ($n = 8$), for the resting, activated and desensitized states, respectively. The pipette solution was Na_o and the intracellular solution was K_o supplemented with 5 $\text{CaK}_2\text{EGTA} + \text{H}_2\text{K}_2\text{EGTA}$ to obtain 0.1 or 1 μM free Ca^{2+} .

Patch-clamp experiments showed that elevating bath Ca^{2+} to 5 mM evoked pronounced, but transient, increases in PCL channel activity in cell-attached patches, although the single-channel conductance remained nearly unchanged (Fig. 3a). This effect of bath Ca^{2+} indicates that an increase in intracellular Ca^{2+} (Ca_i^{2+}) concentration may have triggered the activation. Moreover, during desensitization of the channel, the presence of bath Ca^{2+} resulted in lower single-channel activity than in the resting state (Fig. 3a, points 3 versus 1), consistent with whole-cell observations (Fig. 2a, points 3 versus 1). Direct activation of PCL channel activity by Ca_i^{2+} was shown in inside-out patches (Fig. 3b). Sustained high Ca_i^{2+} (1–10 μM) did not appear to prevent subsequent decreases in channel activity, although the time needed to reach this state varied.

To determine whether an increase in extracellular Ca^{2+} alone could also activate PCL channel activity, oocytes were injected with 50 nl of 50 mM EGTA, a Ca^{2+} chelator, at least 2 h before experiments. Currents due to addition of 5 mM Ca_o^{2+} were much lower than without EGTA and could be repeatedly generated (compare Figs 4a and 2a), indicating that elevating Ca_o^{2+} alone did not trigger activation or desensitization of the PCL channel. It remains to be established whether Ca^{2+} -induced activation and the ensuing desensitization of the channel result from effects of Ca_i^{2+} on the putative

EF-hand within the carboxy terminus (Fig. 1a) or calmodulin¹⁰, or from other mechanisms^{13,14}.

PCL is permeable to Ba^{2+} and Sr^{2+} , but barely permeable to Mg^{2+} (Fig. 4a). Ba^{2+} , Sr^{2+} and Ca^{2+} (80 mM) generated comparable whole-cell inward currents to that generated by 100 mM Na^+ in EGTA-injected oocytes (data not shown). The PCL single-channel inward conductance was 120–135 pS with 80 mM Ba^{2+} , Sr^{2+} or Ca^{2+} . However, its permeation to mono- and divalent cations is not additive, as divalent cations inhibited currents generated by Na^+ . Whereas Ca_o^{2+} inhibited Na^+ currents over a wide voltage range (Fig. 4b), Mg^{2+} inhibited Na^+ currents, by reducing the PCL channel open probability, only at negative potentials (data not shown). Reciprocally, Na^+ inhibited Ca^{2+} influx (Figs 2d and 4b, right panels). Thus, competitive binding may occur between mono- and divalent cations, but not between permeant monovalent cations, as mixing Na^+ and K^+ in different ratios did not reduce PCL conductance. La^{3+} , Gd^{3+} (0.1 mM but not 20 μM) and flufenamate (0.5 mM but not 50 μM), which are inhibitors of nonselective cation channels^{15,16}, exhibited significant inhibitory effects (not shown). Protons also affect PCL channel activity, as low external pH decreased both its resting conductance and Ca^{2+} -activated currents (Fig. 4c).

Ca^{2+} entry was not significantly enhanced by treating oocytes expressing PCL with 1 μM thapsigargin for 2 h (ref. 17), indicating that PCL, unlike members of the TRP family, is not a store-operated capacitative calcium-entry channel. PCL channel activity appears not to be regulated by phosphorylation dependent on protein kinase A (PKA) or protein kinase C (PKC) as neither cyclic AMP nor phorbol myristate acetate significantly affected Ca^{2+} entry (data not shown).

This study shows that PCL is a Ca^{2+} -modulated, Ca^{2+} -permeable, nonselective cation channel. Its ion selectivity, large single-channel conductance and relatively long open time distinguish it from structurally related channels of the TRP family, voltage-gated Ca^{2+} and Na^+ channels^{8–10,18,19} and known endogenous cation channels in *Xenopus* oocytes, including the stretch-activated cation channel and hyperpolarization-activated cation channel^{20–22}. It is unlikely that the unique channel activity observed in PCL-expressing oocytes results from upregulation or modulation of an endogenous channel, as actinomycin D had no effects on the activity (data not shown).

PCL and polycystin-2 share structural features and it is likely that both possess channel properties. We hypothesize that alteration of these channels leads to the abnormal fluid secretion²³ and cellular proliferation that are hallmarks of polycystic kidney disease (PKD). PCL is not yet known to be mutated in PKD but its murine orthologue is removed by a large deletion in *Krd* mice that exhibit renal cysts and other defects⁶. Polycystins-1 and -2 bind each other^{24,25}, and it is possible that PCL also oligomerizes with another polycystin *in vivo*. Elucidating the channel properties of polycystins should provide new therapeutic strategies for PKD. □

Methods

Oocytic expression.

Capped synthetic RNA of human PCL⁶ was synthesized by *in vitro* transcription from full-length cDNA in pTLN2 and injected (25–50 ng per oocyte) into *Xenopus laevis* oocytes prepared as described²⁶. Equal volumes of H_2O were injected into control oocytes. Experiments were performed 1.5–4 days after injection.

Solutions.

Standard Barth's (or NaCl-containing) solution contained (in mM): 100 NaCl, 2 KCl, 1 MgCl_2 , 10 Hepes, pH 7.5. When 100 mM NaCl was replaced with equimolar amounts of other salts, the resulting solutions were named accordingly; for example, Na-glutamate-substituted solution. ' Na_o ', ' K_o ' and ' K140_o ' indicate, respectively, extracellular (or pipette, in cell-attached and inside-out modes) solutions (100 NaCl, 2 KCl, 10 Hepes, pH 7.5; 100 KCl, 10 Hepes, pH 7.5; and 105 K-glutamate, 30 K-fluoride, 5 KCl, 5 EGTA, 5 Hepes, pH 7.3). ' Na_i ', ' K_i ' and ' NMDG_i ' indicate, respectively, intracellular (or pipette, in outside-out mode) solutions (100 NaCl, 2 KCl, 10 Hepes, pH 7.5; 100 KCl, 10 Hepes, pH 7.5; and 100 NMDG-Cl, 10 Hepes, pH 7.5). Cl⁻-free solution for oocyte pre-incubation contained: 80 Na-Hepes, 2 K-gluconate, 1 Mg-gluconate, 0.75 Ca-gluconate, and pH 7.5.

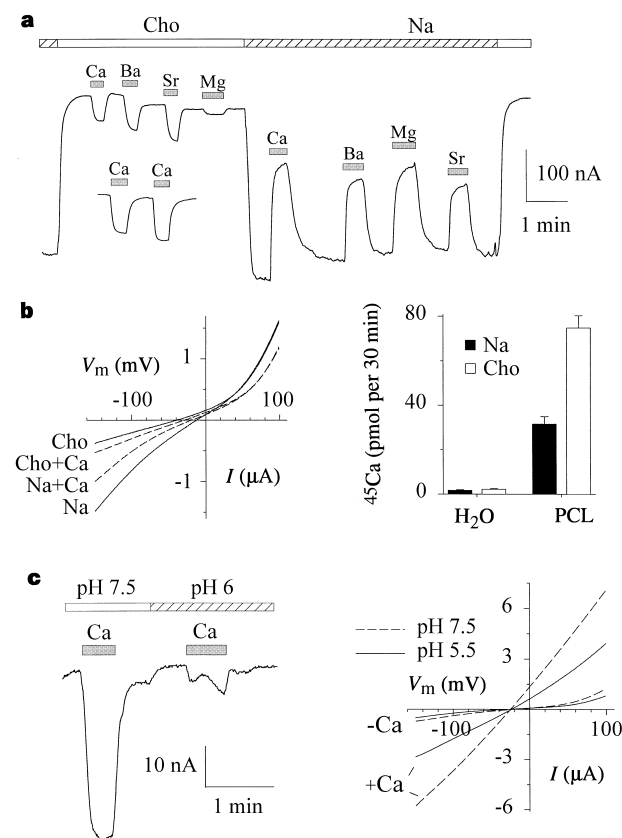


Figure 4 Effects of EGTA, divalent cations and protons. **a**, Currents recorded at -50 mV with an EGTA-injected oocyte in choline-Cl-substituted or NaCl-containing solution. External application of various divalent cations (5 mM) is indicated by grey bars. Inset, currents recorded with another oocyte in choline-Cl-substituted solution. **b**, Left, I - V relations obtained using EGTA-injected oocytes expressing PCL, before or after Ca^{2+} addition to choline-Cl-substituted or NaCl-containing solution. Right, $^{45}\text{Ca}^{2+}$ (1 mM) uptake using H_2O -injected or PCL-expressing oocytes, in NaCl-containing or choline-Cl-substituted solution. **c**, Left, currents recorded at -50 mV with EGTA-injected oocytes in NaCl-containing solution at indicated pH. Right, I - V curves obtained before or after adding 5 mM Ca^{2+} in NaCl-containing solutions in the same oocyte without EGTA injection.

Electrophysiology.

Two-microelectrode voltage-clamp experiments were performed as described²⁷. In experiments involving voltage ramping or holding, currents and voltages were digitized at 0.3 or 200 ms per sample and Bessel filtered at 10 or 0.02 kHz, respectively. We calculated ratios of permeability coefficients for monovalent cations using the equation derived from the GHK equation²⁸: $P_X/P_{Na} = \exp(\Delta V_r/58.5)$, where ΔV_r (mV) is the change in reversal potential. To estimate the permeabilities of divalent cations relative to that of Na^+ under physiological conditions, we compared cation-evoked currents at -50 mV. Experimental results are expressed as mean $\pm$ s.e.m. (n). Patch-clamp experiments were performed as described²⁹. We used a pipette-tip resistance of 5–10 M Ω and seal resistance of >10 G Ω . Single-channel currents were measured with an integrating patch-clamp amplifier, digitized at 0.15 ms per sample and filtered at 3 kHz through an 8-pole Bessel filter. To prevent possible run-down, cAMP, GTP- γ -S and ATP (0.1 mM) were added to intracellular solutions in most cases.

⁴⁵Ca-uptake measurements.

⁴⁵CaCl₂ of 30 and 60 μ M, respectively, was added to uptake solutions containing 1 and 5 mM non-radioactive Ca^{2+} . 8–10 oocytes were incubated in 0.5 ml of uptake solution and uptake was terminated by washing oocytes in ice-cold NaCl-containing solution (pH 7.5). For experiments involving voltage-clamped ⁴⁵Ca uptake, Ca^{2+} -evoked currents and uptake of ⁴⁵Ca were simultaneously measured³⁰ at -50 mV.

Received 21 May; accepted 9 July 1999.

- Gabow, P. A., Autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* **329**, 332–342 (1993).
- The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* **77**, 881–894 (1994).
- The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell* **81**, 289–298 (1995).
- Hughes, J. *et al.* The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nature Genet.* **10**, 151–160 (1995).
- Mochizuki, T. *et al.* PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339–1342 (1996).
- Nomura, H. *et al.* Identification of PKDL, a novel polycystic kidney disease 2-like gene whose murine homologue is deleted in mice with kidney and retinal defects. *J. Biol. Chem.* **273**, 25967–24973 (1998).
- Wu, G. *et al.* Identification of PDK2L, a human PKD2-related gene: tissue-specific expression and mapping to chromosome 10q25. *Genomics* **54**, 564–568 (1998).
- Kiselyov, K. *et al.* Functional interaction between Insp3 receptors and store-operated Htrp3 channels. *Nature* **396**, 478–482 (1998).
- Perez-Reyes, E. *et al.* Molecular characterization of a neuronal low-voltage-activated T-type calcium channel. *Nature* **391**, 896–900 (1998).
- Xia, X. M. *et al.* Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395**, 503–507 (1998).
- Jorgensen, A. J., Bennekou, P., Eskesen, K. & Kristensen, B. I. Annexins from Ehrlich ascites cells inhibit the calcium-activated chloride current in *Xenopus laevis* oocytes. *Pflügers Arch.* **434**, 261–266 (1997).
- Boton, R., Dascal, N., Gillo, B. & Lass, Y. Two calcium-activated chloride conductances in *Xenopus laevis* oocytes permeabilized with the ionophore A23187. *J. Physiol. (Lond.)* **408**, 511–534 (1989).
- Zitt, C. *et al.* Expression of TRPC3 in Chinese hamster ovary cells results in calcium-activated cation currents not related to store depletion. *J. Cell Biol.* **138**, 1333–1341 (1997).
- Zuhlke, R. D. & Reuter, H. Ca^{2+} -sensitive desensitization of L-type Ca^{2+} channels depends on multiple cytoplasmic amino acid sequences of the α_{1C} subunit. *Proc. Natl Acad. Sci. USA* **95**, 3287–3294 (1998).
- Gogelein, H., Dahlem, D., Englert, H. C. & Lang, H. J. Flufenamic acid, mefenamic acid and niflumic acid inhibit single nonselective cation channels in the rat exocrine pancreas. *FEBS Lett.* **268**, 79–82 (1990).
- Kunze, D. L., Sinkins, W. G., Vaca, L. & Schilling, W. P. Properties of single *Drosophila* Trp1 channels expressed in Sf9 insect cells. *Am. J. Physiol.* **272**, C27–C34 (1997).
- Gillo, B. *et al.* Coexpression of *Drosophila* TRP and TRP-like proteins in *Xenopus* oocytes reconstitutes capacitative Ca^{2+} entry. *Proc. Natl Acad. Sci. USA* **93**, 14146–14151 (1996).
- Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Birnbaumer, L. *et al.* On the molecular basis and regulation of cellular capacitative calcium entry: roles for TRP proteins. *Proc. Natl Acad. Sci. USA* **93**, 15195–15202 (1996).
- Yang, X. C. & Sachs, F. Block of stretch-activated ion channels in *Xenopus* oocytes by gadolinium and calcium ions. *Science* **243**, 1068–1071 (1989).
- Lane, J. W., McBride, D. W. Jr & Hamill, O. P. Ionic effects on amiloride block of the mechanosensitive channel in *Xenopus* oocytes. *Br. J. Pharmacol.* **108**, 116–119 (1993).
- Tzounopoulos, T., Maylie, J. & Adelman, J. P. Induction of endogenous channels by high levels of heterologous membrane proteins in *Xenopus* oocytes. *Biophys. J.* **69**, 904–908 (1995).
- Sullivan, L. P., Wallace, D. P. & Grantham, J. J. Chloride and fluid secretion in polycystic kidney disease. *J. Am. Soc. Nephrol.* **9**, 903–916 (1998).
- Tsiokas, L., Kim, E., Arnould, T., Sukhatme, V. P. & Walz, G. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc. Natl Acad. Sci. USA* **94**, 6965–6970 (1997).
- Qian, F. *et al.* PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nature Genet.* **16**, 179–183 (1997).
- Saadi, I. *et al.* Molecular genetics of cystinuria: mutation analysis of SLC3A1 and evidence for another gene in type I (silent) phenotype. *Kidney Int.* **54**, 48–55 (1998).
- Chen, X.-Z., Shayakul, C., Berger, U. V., Tian, W. & Hediger, M. A. Characterization of a rat Na^+ -dicarboxylate cotransporter. *J. Biol. Chem.* **273**, 20972–20981 (1998).
- Hille, B. in *Ionic Channels of Excitable Membranes* (ed. Hille, B.) 105–108 (Sinauer, Sunderland, Massachusetts, 1992).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).

30. Chen, X.-Z., Zhu, T., Smith, D. E. & Hediger, M. A. Stoichiometry and kinetics of the rat high-affinity H^+ -coupled peptide transporter PepT2. *J. Biol. Chem.* **274**, 2773–2779 (1999).

Acknowledgements

We thank P. Fong for providing pTLN2. X.-Z.C. is a recipient of the International Human Frontier Science Program Long-Term Fellowship. This work is supported by NARSAD and the Stanley Foundation (P.M.V.), the St. Giles Foundation (E.M.B.) and NIH (S.T.R., E.M.B., M.A.H. and J.Z.).

Correspondence and requests for materials should be addressed to J.Z. (e-mail: zhou@rics.bwh.harvard.edu) or M.A.H. (e-mail: mhediger@rics.bwh.harvard.edu).

A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*

Maureen M. Barr & Paul W. Sternberg

Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

The stereotyped mating behaviour of the *Caenorhabditis elegans* male is made up of several substeps: response, backing, turning, vulva location, spicule insertion and sperm transfer. The complexity of this behaviour is reflected in the sexually dimorphic anatomy and nervous system¹. Behavioural functions have been assigned to most of the male-specific sensory neurons by means of cell ablations; for example, the hook sensory neurons HOA and HOB are specifically required for vulva location². We have investigated how sensory perception of the hermaphrodite by the *C. elegans* male controls mating behaviours. Here we identify a gene, *lov-1* (for location of vulva), that is required for two male sensory behaviours: response and vulva location. *lov-1* encodes a putative membrane protein with a mucin-like, serine–threonine-rich amino terminus³ followed by two blocks of homology to human polycystins, products of the autosomal dominant polycystic kidney-disease loci PKD1 and PKD2 (ref 4). *LOV-1* is the closest *C. elegans* homologue of PKD1. *lov-1* is expressed in adult males in sensory neurons of the rays, hook and head, which mediate response, vulva location, and potentially chemotaxis to hermaphrodites, respectively^{2,5}. PKD-2, the *C. elegans* homologue of PKD2, is localized to the same neurons as *LOV-1*, suggesting that they function in the same pathway.

We examined the mating behaviour of existing mutants that are defective in sensory behaviours including mechanosensation, osmotic avoidance and chemotaxis to soluble and volatile odorants. Only males with severe defects in all sensory neuron cilia (*daf-10*, *osm-5*, *osm-6* and *che-3*) were defective of vulva location (Table 1); all cilia in *C. elegans* are in the dendritic endings of sensory neurons^{5,6}. Only ciliated neurons express *osm-6::gfp*, with male-specific expression in four CEM head neurons and neurons of the rays and copulatory spicules⁷. Expression of *osm-6::gfp* begins at the L4 stage in neuronal cell bodies and extends to dendrites as neuronal outgrowth proceeds. The RnA and RnB neurons of each ray (rays 1–9), both HOA and HOB hook neurons, the spicule neurons SPV and SPD, and the PCB postcloacal sensilla neurons accumulate green fluorescent protein (GFP). *OSM-6* may be required for the structure and function of ciliated neurons in the tail of the adult male, just as it is for neurons involved in many sensory behaviours^{8,9}.

By screening for mutants defective in vulva location, we identified *lov-1(sy552)*, which results in specific response and vulva-location

defects. When a *lov-1(+)* male encounters a hermaphrodite, it responds by placing its tail flush on the hermaphrodite, then backs along her body, and turns at her ends until it stops at her vulva. *lov-1* mutant males frequently fail to respond to hermaphrodite contact. Compared with the 88% responsiveness of wild-type males ($n = 82$), *sy552* male responsiveness is 24% ($n = 37$, $P < 0.0001$). When responsive is initiated, *lov-1(sy552)* mutants back and turn normally but frequently pass the vulva (Table 1), although spicule insertion and sperm transfer are unaffected. *lov-1(sy552)* males exhibit high mating efficiency with severely paralysed *unc-52* hermaphrodites, but sire few progeny with active, moving *dpy-17* hermaphrodites, presumably because a paralysed partner is an easy target, whereas an active partner evades the *lov-1* mutant male. *lov-1(sy552)* mutants appear normal for movement, egg-laying, nose touch, tap, mechanosensation and osmotic avoidance.

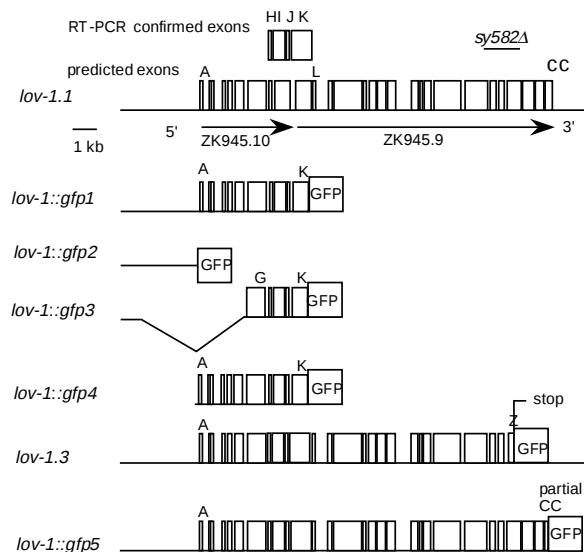
We cloned the *lov-1* gene on the basis of its genetic map position (Fig. 1). *sy552* is recessive. *mnDf21/sy552* and *sy552/sy552* males are phenotypically indistinguishable, so it is likely that *sy552* decreases the function of *lov-1*. *lov-1* falls between the breakpoints of *eDf21* and *mnDf21*. Cosmids in this region were injected and only ZK945 rescued *lov-1* defects (4 out of 5 stable lines). A 16.9-kilobase (kb) subclone (plov-1.1) rescued both the response and vulva-location defects of *sy552*, but shorter subclones did not (Fig. 1). Both a 6.7-kb (plov-1::gfp1) and a plov-1.1 frameshift clone (plov-1.3) failed to rescue *sy552* defects, yet dominantly inhibited vulva location but not response (Fig. 1).

lov-1 encodes a predicted 3,125-amino-acid membrane-bound protein with a serine-threonine-rich, potential extracellular

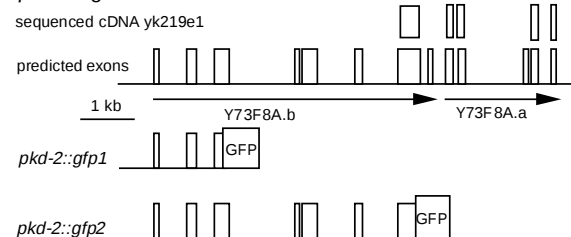
domain similar to mucins³, a polycystin homology block 1 (26% identity), and a carboxy-terminal polycystin block 2 with 20% identity to polycystin proteins 1, 2 and 2L, encoded by the PKD1, PKD2 and PKDL loci, respectively (Fig. 2). A Kyte-Doolittle hydropathy plot predicts several transmembrane domains. No signal peptide is predicted in LOV-1. Mutations in PKD1 or PKD2 account for 95% of autosomal dominant polycystic disease⁴. There is similarity between LOV-1, the polycystins and the TRP and voltage-activated calcium and potassium family of channels in the transmembrane spanning region^{10–14}. LOV-1 lacks the Ca²⁺-binding EF-hand of polycystins 2 and L, the characteristic pore region of channels, and a cytoplasmic coiled-coil tail of all three polycystins (Fig. 2), which mediate hetero- and homotypic interactions between polycystin 1 and polycystin 2 (refs 15, 16). However, truncation of 58 C-terminal amino acids of LOV-1 (LOV-1::GFP5) destroys its function (Fig. 1), indicating that the LOV-1 cytoplasmic tail may be important for LOV-1 action. LOV-1 possesses a potential nucleotide-binding domain (Fig. 2) not present in the human polycystins. Polycystins have recently been implicated in signal transduction^{17–19}. Consistent with a signalling function, PKD2 and the TRPC1 channel interact *in vitro*, as do PKD1 and TRPC family members²⁰.

We isolated *sy582Δ*, a genomic deletion of *lov-1* that encodes a truncated protein lacking the polycystin/channel homology domain (Fig. 1). *lov-1(sy582Δ)* has a similar genetic and phenotypic profile to *sy552* (Table 1). *sy552/sy582Δ* males are phenotypically indistinguishable from *sy552/Df* males, consistent with *sy582Δ* being a null allele. Thus, the polycystin block 2 is required for LOV-1 activity.

lov-1 genomic structure and GFP fusions



pkd-2 genomic structure and GFP fusions



	Behavioural phenotypes	Expression pattern	Subcellular localization
	<i>lov-1(sy552)</i> rescue	wild type	
<i>lov-1.1</i>	+ (6/8)	+	
<i>lov-1::gfp1</i>	- (0/2)	DN	male-specific sensory neurons: CEMs, HOB, rays
<i>lov-1::gfp2</i>		+	faint, nonspecific
<i>lov-1::gfp3</i>		+	faint, male-specific: 1 or 2 unid'ed neurons
<i>lov-1::gfp4</i>		+	faint, nonspecific
<i>lov-1.3</i>	- (0/2)	DN	not expressed
<i>lov-1::gfp5</i>	- (0/3)	+	male-specific sensory neurons: CEMs, HOB, rays
			Cell body only
<i>pkd-2::gfp1</i>	-	+	male-specific sensory neurons: CEMs, HOB, rays nonspecific, ring neurons (weak)
<i>pkd-2::gfp2</i>	-	+	male-specific sensory neurons: CEMs, HOB, rays
			Uniform (cell body, axon, dendrite)
			Basodendritic (Cell body, sensory ending)

Figure 1 *lov-1* and *pkd-2* genomic structures, constructs, rescue data and expression patterns. The line above the *lov-1* gene indicates the 1,055-bp deletion in *lov-1(sy582Δ)*.

Numbers in parentheses indicate the ratio of rescuing stable lines to the number of total stable lines examined. DN, dominant negative.

We examined the expression pattern of LOV-1::GFP reporters. A 6.7-kb fusion (LOV-1::GFP1) and a 15.4-kb fusion lacking only 58 amino acids (LOV-1::GFP5) direct expression in male-specific sensory neurons (Figs 1 and 3). Shorter versions of *lov-1::gfp1* are not expressed in the same set of male-specific neurons, nor do they act as dominant negatives (Fig. 1). *lov-1::gfp1* and *lov-1::gfp5* are expressed in male-specific neurons, specifically the CEM neurons, the hook neuron HOB, and the sensory ray neurons. Expression begins during late L4 lethargus and peaks in the adult male. LOV-1::GFP1 is localized at high levels in the cell body (punctate cytoplasmic expression) and ciliated endings (Fig. 3), whereas LOV-1::GFP5 is found exclusively in cell bodies. The temporal and spatial regulation of *lov-1* is concordant with its role in adult male mating behaviour. Rays mediate responses to contact with a hermaphrodite², the hook mediates vulva location², and the CEMs may be involved in chemosensation⁵.

The functions of the polycystins and the molecular basis of kidney cystogenesis are not known, but polycystin 1 and polycystin 2 have been proposed to function in concert⁴. The *C. elegans* genome contains a PKD2 homologue with 27% overall identity (Fig. 2). Like PKD2, PKD-2 possesses six membrane-spanning domains, a positively charged fourth membrane-spanning segment, a pore region and the coiled-coil domain of all polycystins. PKD-2 is localized to the same male-specific sensory neurons as LOV-1 (Figs 1 and 3). Rarely, very faint non-sex- and non-stage-specific expression of *pkd-2::gfp1* is observed in nerve-ring cell bodies. Expression of *pkd-2::gfp1* (which lacks transmembrane domains) is uniform throughout the HOB, ray and CEM neurons. In contrast, the basodendritic localization of PKD-2::GFP2 (which contains membrane-spanning regions) is identical to that of LOV-1::GFP1 (Figs 1 and 3). If the expression patterns of *lov-1::gfp1* and *pkd-2::gfp2* do indicate protein localization, then the basodendritic localization of LOV-1 and PKD-2 is consistent with plasma-membrane localization and a role for polycystins in maintaining cell polarity²¹. Although GFP localization might not reflect protein localization, and subcellular protein localization remains ambiguous, LOV-1 and PKD-2 are colocalized to male-specific sensory neurons and so might function in the same pathway.

Although we cannot rule out a subtle developmental defect, the following lines of evidence suggest that LOV-1 has a sensory

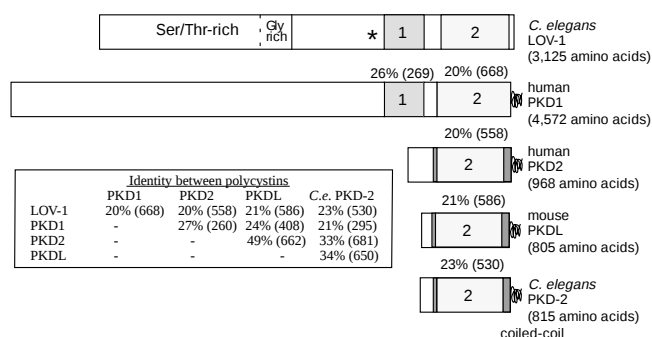


Figure 2 LOV-1 structural features and homologies. The LOV-1 N terminus is serine/threonine rich with several potential glycosylation sites followed by an ATP/GTP-binding domain (asterisk) and two polycystin blocks of homology. Block 2 also shows homology with voltage-activated Ca²⁺-channel and TRP-channel families. The number of identical amino acids between LOV-1 and a particular polycystin is indicated. A coiled-coil is predicted in the middle of LOV-1, using the most stringent criteria for the COILS program. PKD-2 was identified in a BLAST search of unpublished sequences available through the Sanger Centre. A partial corresponding cDNA, yk219e1, was sequenced and aligned to the Genefinder ORFs Y73F8a.b and Y73F8a.a, changing the predicted size of PKD-2 from 913 to 815 amino acids. PKD-2 is more similar to PKD2 (33% identity, 52% similarity, 8% gaps over 681 amino acids) than LOV-1 (22% identity, 42% similarity, 13% gaps over 530 amino acids). Additional similarity between PKD2, PKD2L and PKD-2 is denoted by a hatching. Overall identity between polycystins, LOV-1 and PKD-2 is indicated.

Table 1 Vulva-location behaviour of wild-type and mutant males

Genotype	Vulva location efficiency (%)	Significantly different from wild-type (P-value)	Number of males
<i>him-5</i> (wild type)	96		101
<i>osm-1(e1803)</i>	65	No (0.0738)	6
<i>daf-10(p821)</i>	48	Yes (0.0004)	7
<i>osm-5(p813); him-5</i>	26	Yes (0.0002)	5
<i>osm-6(p811)</i>	32	Yes (0.0003)	7
<i>che-3(e1124)</i>	69	Yes (0.02666)	4
<i>lov-1(sy582Δ)</i>	33	Yes (<0.0001)	11
<i>lov-1(sy552); him-5</i>	30	Yes (<0.0001)	73

lov-1(sy552); him-5(e1490), *lov-1(sy582Δ)* and all cilia-defective mutants were also response-defective. Males that eventually responded were scored for vulva-location behaviour. Males were observed for a minimum of 10 vulva encounters or until spicule insertion, whichever occurred first. Mann-Whitney tests determined P-values. The following mutants were also examined and found to be normal for response and vulva location: *osm-3(e1806)*; *him-5*, *osm-7(n1515)*, *osm-8(n1518)*, *osm-10(n1602)*, *osm-11(n1604)*, *osm-12(n1606)*, *mec-3(e1338)* *him-8(e1489)*, *mec-4(e1611)*, *mec-5(e1340)*, *mec-7(n434)*, *mec-7(e343)*, *mec-8(e398)*, *mec-9(e1494)*, *che-1(e1034)*, *che-12(e1812)*, *odr-1(n1936)*, *odr-2(n2145)*, *odr-3(n2150)*, *odr-4(n2144ts)*, *odr-5(ky9)*, *odr-6(ky1)*, *odr-7(ky4)*, *odr-10(ky32)* and *daf-11(m47ts)*.

function. The rays of many cilium-structure mutants ectopically fill with FITC⁹ but this defect is not observed in *lov-1(sy552)*. By using a functional OSM-6::GFP fusion protein to image sensory neurons, we found that *lov-1* sensory endings appear structurally normal. Also, LOV-1::GFP1 and PKD-2::GFP2 are not mislocalized in *lov-1* mutants suggesting that their neuronal structure is intact. Expression of *lov-1::gfp* peaks in the adult male, as opposed to expression of *osm-6::gfp*, which begins about the time of neuronal outgrowth. Furthermore, the CEMs are formed during embryogenesis, whereas expression of *lov-1::gfp* in the CEMs is strictly restricted to L4 lethargus and adulthood. Finally, scanning electron microscope and Normarski images show that hook and ray structures are normal.

Chemosensation and mechanosensation are probably involved in vulva location (M.B. and P.S., manuscript in preparation). Because mechanosensory mutant males are not defective in response or vulva location, these behaviours may be mediated by a set of gene products other than those involved in body-touch mechanosensation. LOV-1 might act in a sensory signalling pathway, coupling voltage-activated signalling or store-operated conductance in a similar way to hair-cell mechanosensation²² and touch response in *C. elegans*²³ or light-induced cation conductance in *Drosophila* photoreceptor cells²⁴, respectively. Alternatively, LOV-1 might function as a molecular scaffold for other molecules, such as PKD-2, in establishing or maintaining neuronal cell polarity, or ultrastructurally in the assembly of male-specific cilia.

Methods

Mating efficiency and mating behaviour

Standard assays were performed using hermaphrodite strains N2, *unc-52 (e444)*, *dpy-17(e164)*, and *unc-31(e169)* with *him-5(1490)* or heat-shock-generated males^{2,25,26}. Responsiveness reflects the percentage of males successfully responding to hermaphrodite contact within 10 min. An individual male's vulva-location ability was calculated as the number of positive vulva locations divided by the total number of vulva encounters. Vulva-location efficiency indicates the average behaviour of a genotypic population. Pairwise comparisons were made using Mann-Whitney nonparametric and two-sided *t*-tests.

Genetic screen for and mapping of *lov-1*

An F₂ clonal screen of EMS mutagenized PS1395 hermaphrodites [*plg-1(e2001d)*; *him-51*²⁷ (K. Liu and P.S., unpublished) identified *lov-1(sy552)*. Mapping was performed using standard methods. From *unc-4(e120)* *let-25(mn25)* crosses, the *sy552* genotype was screened by complementation. Of 12 *Unc* non-Let recombinants, 2 segregate the *lov-1* mutant phenotype.

Transformation rescue of *lov-1(sy552)* mutants

Cosmids and plasmids (15–100 ng μl⁻¹) in the region from the right breakpoint of *eDf21* to the right breakpoint of *mnDf21* and PHA-1 (100 ng μl⁻¹ pBX) was injected into *lov-1(sy552)*; *pha-1(e2123ts)*; *him-5*. Stable lines were selected at either 19 or 25 °C (ref. 28). A 16.9-kb *Hind*III fragment of ZK945 was cloned into pBS(SK+) (plov1.1). A frameshift in plov1.1 was created at nucleotide 17724 of ZK945 by inserting a *Bss*HII GFP fragment

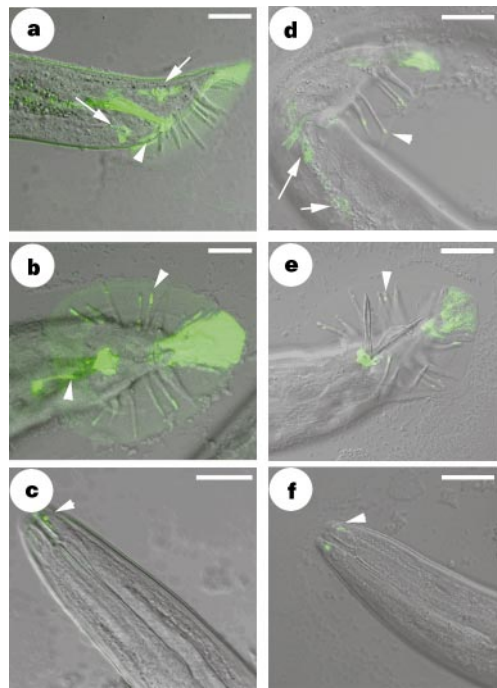


Figure 3 LOV-1::GFP1 and PKD-2::GFP2 are colocalized to adult male sensory-neuron cell bodies and dendrites. The spicules, hook structure and posteriormost fan region autofluoresce. Arrows, neuronal cell bodies; arrowheads, dendrites or ciliated endings. Images (merged DIC and fluorescence) were obtained using confocal microscopy. **a–c**, *lov-1::gfp1*. **a**, HOB and ray cell bodies (arrows), HOB dendritic process (arrowhead). **b**, HOB and ray process 5 (arrowheads). **c**, Ciliated endings in nose tip from male-specific cephalic CEM neurons (cell bodies not shown). **d–f**, *pkd-1::gfp2*. **d**, Ray cell bodies (arrow) and ray process 2 (arrowhead). **e**, Ray process 5 (arrowhead). **f**, Male-specific cephalic CEM ciliated endings (arrow). Scale bar, 20 μ m.

from plasmid pPD95.02 (A. Fire, personal communication) out of frame into the *StuI* site of plov-1.1, creating plov-1.3.

PCR screen for genomic deletion of *lov-1*

We followed the procedure of G. Moulder and R. Barstead (personal communication) and screened approximately 315,000 haploid genomes by using primers designed to delete the PKD/channel domain. *lov-1*(*sy582Δ*) deletes nucleotides 16972–18027 of ZK945.

DNA-sequence analysis

DNA sequence analysis of cDNA clones generated by RT-PCR from *him-5*(*e1490*) RNA revealed three exons in the junction between ZK945.10 and ZK945.9: one from 25742–25195, a second from 25151–25071, and a third initiating at position 25021, corresponding to exons I, J and K, in Fig. 1, respectively. Sequencing of the partial *pkd-2* cDNA *yk219e* indicates coding regions of Y73F8A.b and Y73F8A.a at 4932–4480, 3813–3562, 3400–3147, 2100–1972, and 1692–1606.

Expression analysis

GFP²⁹ was used as a marker for *lov-1* and *pkd-2* gene expression. plov-1::GFP1 was constructed by cloning a 6.7-kb *HindIII*–*BamHI* fragment of plov-1.1 into the vector pPD95.81, plov-1::GFP2 a *HindIII*–*HpaI* fragment. plov-1::GFP3 and plov-1::GFP4 are *SacI* and *HindIII*–*HpaI* (Klenow filled-in and relegated) deletions of plov-1::GFP1, respectively. plov-1::gfp5 was constructed by cloning a 15.4-kb *HindIII*–*AfeI* fragment of plov-1.1 into the *HindIII*–*SmaI* site of pPD95.79. ppkd-2.1, ppkd-2::gfp1 and ppkd-2::gfp2 were constructed by cloning PCR-amplified 8.9-kb, 2.0-kb and 5.9-kb fragments into the vectors pPD.95.97, pPD95.75 and pPD95.77, respectively. Cells were identified by comparing Nomarski and fluorescent or confocal images of the same animals to determine cell-body position³⁰. HOB assignment was confirmed by laser ablation of precursor cells.

Received 23 April; accepted 16 July 1999.

- Hodgkin, J. in *The Nematode C. elegans* (ed. Wood, B.) 243–279 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Liu, K. S. & Sternberg, P. W. Sensory regulation of male mating behavior in *Caenorhabditis elegans*. *Neuron* **14**, 79–89 (1995).
- Carraway, K. L. & Fregien, N. Mucin structure and function: Insights from molecular biology. *Trends Glycosci. Glycotechnol.* **7**, 31–44 (1995).
- Torres, V. D. New insights into polycystic kidney disease and its treatment. *Curr. Opin. Nephrol. Hypertens.* **7**, 159–169 (1998).
- Ward, S., Thomson, N., White, J. G. & Brenner, S. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* **160**, 313–337 (1975).
- White, J. G., Southgate, D., Thomson, J. N. & Brenner, S. The structure of the nervous system of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Collet, J., Spike, C. A., Lundquist, E. A., Shaw, J. E. & Herman, R. K. Analysis of *osm-6*, a gene that affects sensory cilium structure and sensory neuron function in *Caenorhabditis elegans*. *Genetics* **148**, 187–200 (1998).
- Kaplan, J. M. & Horvitz, H. R. A dual mechanosensory and chemosensory neuron in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 2227–2231 (1993).
- Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).

- Mochizuki, T. *et al.* PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **242**, 1339–1342 (1996).
- Montell, C. & Rubin, G. M. Molecular characterization of the *Drosophila trp* locus: a putative integral membrane protein required for phototransduction. *Neuron* **2**, 1313–1323 (1989).
- Sanford, R. *et al.* Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. *Hum. Mol. Genet.* **9**, 1483–1489 (1997).
- Nomura, H. *et al.* Identification of PKDL, a novel polycystic kidney disease 2-like gene whose murine homologue is deleted in mice with kidney and retinal defects. *J. Biol. Chem.* **273**, 25967–25973 (1998).
- Wu, G. *et al.* Identification of PKD2L, a human PKD2-related gene: Tissue-specific expression and mapping to chromosome 10q25. *Genomics* **54**, 564–568 (1998).
- Qian, F. *et al.* PKD1 interacts with PKD2 through a probably coiled-coil domain. *Nature Genet.* **16**, 179–183 (1997).
- Tsiokas, L., Kim, E., Arnould, T., Sukhatme, V. P. & Walz, G. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc. Natl Acad. Sci. USA* **94**, 6965–6970 (1997).
- Arnould, T. *et al.* The polycystic kidney disease 1 gene product mediates protein kinase C α -dependent and c-Jun N-terminal kinase-dependent activation of the transcription factor AP-1. *J. Biol. Chem.* **273**, 6013–6018 (1998).
- Parnell, S. C. *et al.* The polycystic kidney disease-1 protein, polycystin-1, binds and activates heterotrimeric G-proteins *in vitro*. *Biochem. Biophys. Res. Commun.* **251**, 625–631 (1998).
- Kim, E. *et al.* The polycystic kidney disease 1 gene product modulates Wnt signaling. *J. Biol. Chem.* **274**, 4947–4953 (1999).
- Tsiokas, L. *et al.* Specific association of the gene product of PKD2 with the TRPC1 channel. *Proc. Natl Acad. Sci. USA* **7**, 3934–3939 (1999).
- Sullivan, L. P., Wallace, D. P. & Grantham, J. J. Epithelial transport in polycystic kidney disease. *Physiol. Rev.* **78**, 1165–1191 (1998).
- Hudspeth, A. J. How the ear's works work. *Nature* **341**, 397–404 (1989).
- Driscoll, M. & Kaplan, J. in *C. elegans II* (ed. Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 645–677 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Montell, C. TRP trapped in the fly signalling web. *Curr. Opin. Neurobiol.* **8**, 389–397 (1998).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Hodgkin, J. Male phenotypes and mating efficiency in *Caenorhabditis elegans*. *Genetics* **103**, 43–64 (1983).
- Hodgkin, J. & Doniach, T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* **146**, 149–164 (1997).
- Schnabel, H. & Schnabel, R. An organ-specific differentiation gene, *pha-1*, from *Caenorhabditis elegans*. *Science* **250**, 686–688 (1990).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. D. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
- Sulston, J. D., Albertson, D. G. & Thomson, J. N. The *Caenorhabditis elegans* male: Postembryonic development of nongonadal structures. *Dev. Biol.* **78**, 542–576 (1980).

Acknowledgements

We thank J. Copeland for help isolating *lov-1*(*sy582Δ*); L. Jiang, R. Garcia and R. Ballester for discussions and constructive criticisms; members of our lab, S. Myers and M. Snow for experimental suggestions; D. Sherwood and B. Smith for assistance with confocal microscopy; A. Fire for vectors; and R. Herman for *osm-6::gfp*. The *Caenorhabditis* Genetics stock Center and Sanger Center provided numerous strains, cosmid and sequencing data. This work was supported by the Howard Hughes Medical Institute, with which P.W.S. is an investigator and M.M.B. is an associate, and by the Seaver Institute.

Correspondence and requests for materials should be addressed to P.W.S. (e-mail: pws@cco.caltech.edu).

Dystrophin expression in the *mdx* mouse restored by stem cell transplantation

Emanuela Gussoni*, Yuko Soneoka†‡, Corinne D. Strickland*, Elizabeth A. Buzney*, Mohamed K. Khan§, Alan F. Flint*†, Louis M. Kunkel*†‡ & Richard C. Mulligan†‡

* Division of Genetics, † Howard Hughes Medical Institute and § Department of Surgical Research, Children's Hospital, Boston, Massachusetts 02115 USA

‡ Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

The development of cell or gene therapies for diseases involving cells that are widely distributed throughout the body has been severely hampered by the inability to achieve the disseminated delivery of cells or genes to the affected tissues or organ¹. Here we report the results of bone marrow transplantation studies in the *mdx* mouse, an animal model of Duchenne's muscular dystrophy², which indicate that the intravenous injection of either normal haematopoietic stem cells or a novel population of muscle-derived stem cells into irradiated animals results in the reconstitution of the haematopoietic compartment of the transplanted recipients, the incorporation of donor-derived nuclei into muscle, and the partial restoration of dystrophin expression in the affected muscle. These results suggest that the transplantation of different stem cell populations, using the procedures of bone marrow transplantation, might provide an unanticipated avenue for treating muscular dystrophy as well as other diseases where the systemic delivery of therapeutic cells to sites throughout the body is critical. Our studies also suggest that the inherent developmental potential of stem cells isolated from diverse tissues or organs may be more similar than previously anticipated.

To determine whether transplantation of bone marrow cells carrying a wild-type dystrophin gene could restore dystrophin expression within the myofibres of the *mdx* mouse, nine female *mdx* mice were lethally irradiated and subsequently injected via the tail vein with 5×10^5 or $1-5 \times 10^7$ bone marrow cells from normal male C57BL/10 mice (Table 1). Five weeks after transplantation, the haematopoietic cells of all transplanted animals were donor-derived (of male origin), as determined by fluorescence *in situ* hybridization (FISH) analysis for Y-chromosome-specific sequences (data not shown). At 5, 8 and 12 weeks after bone marrow transplantation, the tibialis anterior muscle from each transplanted animal was analysed for dystrophin expression by immunohistochemistry, followed by FISH analysis using a Y-chromosome-specific probe to detect donor-derived (male) cells (Table 1; Fig. 1). Because previous studies of nuclear domains in myofibres had shown that the expression of a protein can extend several micrometres away from the source nucleus³⁻⁶, around 20 serial sections were analysed to attempt to associate a dystrophin-positive fibre with a donor-derived (male) nucleus.

At five weeks after bone marrow transplantation, less than 1% of the myofibres expressed dystrophin (Table 1), similar to background reversion levels described in other studies⁷. Eight weeks after bone marrow transplantation, however, about 1% of the muscle fibres expressed dystrophin in their normal sarcolemmal location, with 25-63% of dystrophin-expressing fibres containing detectable fused donor-derived nuclei (Table 1). By 12 weeks after bone marrow transplantation, as many as 10% of the muscle fibres within an individual mouse expressed dystrophin, with 10-30% of the dystrophin-positive myofibres containing detectable fused Y-chromosome-positive nuclei (Table 1; Fig. 1a, b).

Although other studies have documented the capacity of bone-marrow-derived cells to give rise to muscle^{8,9}, the identity of the cells with that potential had not been determined. To address this issue, we asked whether highly purified haematopoietic stem cells (HSCs) could give rise to dystrophin-positive myofibres after bone marrow transplantation. HSCs were isolated from the bone marrow of normal male C57BL/10 mice by fluorescence-activated cell sorting (FACS) purification of Hoechst 33342-stained (H0342) cells, as described¹⁰. These HSCs, termed SP cells, are Sca-1⁺, c-Kit⁺, CD43⁺, CD45⁺, lineage marker (B220, Mac-1, Gr-1, CD4, CD5, and CD8)-low/negative and CD34-negative, and can completely engraft recipients at very low cell numbers (100-500 cells per mouse)^{10,11}. We injected 2,000-5,000 bone marrow SP cells of male origin into the tail veins of nine lethally irradiated female *mdx* mice (Table 1). Mice were killed 5, 8 and 12 weeks after stem cell injection and FISH analysis of the bone marrow confirmed that the female host was completely reconstituted with male donor cells (data not shown). The tibialis anterior muscle was examined for dystrophin expression and the presence of male donor cells by the analysis of serial sections, as described above (Table 1). As was the case with recipients engrafted with unfractionated bone marrow cells, less than 1% of the myofibres expressed dystrophin at 5 weeks (Table 1). At 8 weeks, up to 1% of the muscle fibres expressed dystrophin, with 20-40% of dystrophin-expressing fibres containing fused Y-chromosome-positive nuclei (Table 1). At 12 weeks, dystrophin was detected in up to 4% of the myofibres, with 10-30% of these containing fused donor nuclei (Table 1; Fig. 1c-f).

Donor-derived nuclei were associated with both individual dystrophin-positive fibres and clusters of dystrophin-positive

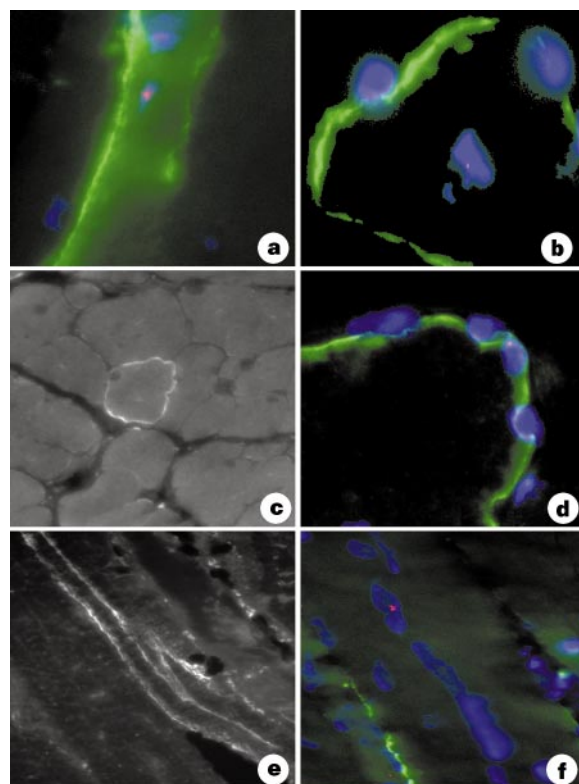


Figure 1 Dystrophin expression and detection of Y-chromosome-positive nuclei in the tibialis anterior muscle 12 weeks after whole bone marrow or haematopoietic stem cell transplantation into lethally irradiated female *mdx* recipients. Donor nuclei (nuclei counterstained with DAPI in blue and Y chromosomes shown as red hybridization signals) were found fused to dystrophin-positive myofibres (in green) at 12 weeks after either whole bone marrow (a, b) or highly purified haematopoietic stem cell transplantation (c-f). In f, the fused donor nucleus was detected further upstream of the dystrophin-positive fibre shown in e.

Table 1 Detection of donor-derived dystrophin-positive muscle fibres after transplantation of whole bone marrow or bone marrow SP cells

No. of animals	Weeks after injection	No. cells injected*	% Dys ⁺ myofibres (total)†	% Y ⁺ nuclei in myofibres‡
3	5	5 × 10 ⁵ per animal (WBM)	0.5; 0.5; 1 (795)	0; 0; 0
3	8	5 × 10 ⁵ ; 5 × 10 ⁷ ; 5 × 10 ⁷ (WBM)	1; 1; 1 (919)	63; 36; 25
3	12	1 × 10 ⁷ ; 5 × 10 ⁷ ; 5 × 10 ⁷ (WBM)	10; 5; 1 (552)	12; 8; 33
4	5	2,000 per animal (BM-SP)	0.5; 0.5; 0.6; 0.7 (1,016)	0; 0; 0; 0
2	8	2,000 per animal (BM-SP)	1; 0.5 (1,715)	39; 22
3	12	2,000; 5,000; 5,000 (BM-SP)	4; 1; 3 (1,214)	12; 28; 14

* WBM: whole bone marrow; BM-SP: bone-marrow SP cells.

† Percentage of dystrophin-positive fibres in each animal. (total) represents the average total number of myofibres.

‡ Percentage of dystrophin-positive fibres for which a Y-chromosome-positive nucleus was found in the skeletal muscle of each animal for all sections analysed (on average, one or none per section).

fibres. Among 39 photographed sections, 4–8 Y⁺ nuclei were centrally located, 5–6 nuclei were peripherally located within myofibres and 7–9 nuclei were clearly fused to host myofibres but their location (central or peripheral) was uncertain. However, in some cases where clusters of dystrophin-positive fibres were analysed, no more than a single male nucleus was found. This finding could be due to either preferential fusion of Y⁺ donor cells to clusters of revertant fibres or, alternatively, to an underestimate of the actual proportion of dystrophin-positive fibres associated with donor-derived nuclei. In fact, only a portion of each myofibre was analysed by immunohistochemistry and FISH, and control

experiments showed that the FISH analysis does not identify 100% of male nuclei. In addition, although the percentage of fibres associated with a donor-derived nucleus appeared to decrease at longer times after transplantation of bone marrow cells (Table 1), this may be simply because the domain of dystrophin expression from an individual donor nucleus increasing with time^{3–6}.

As the Hoechst 33342 staining/FACS method for purification of haematopoietic stem cells appears to depend upon a set of physical properties that are common to haematopoietic stem cells from many species, including humans, in a manner independent of antibodies¹¹, we next asked whether this method could be used to isolate putative stem cells from skeletal muscle. Mononuclear cells were isolated from 3–5-week-old mouse skeletal muscle¹², stained with 12.5 µg ml⁻¹ of H0342, and analysed as described¹⁰. As observed after the staining of bone marrow cells, FACS analysis of muscle cells revealed a side population (SP) displaying low staining with H0342 and a main population (MP) of cells that were more brightly stained with the dye (Fig. 2a). As was the case for bone marrow SP cells, the muscle SP cell population disappeared upon addition of verapamil (Fig. 2b), a drug that blocks the efflux of

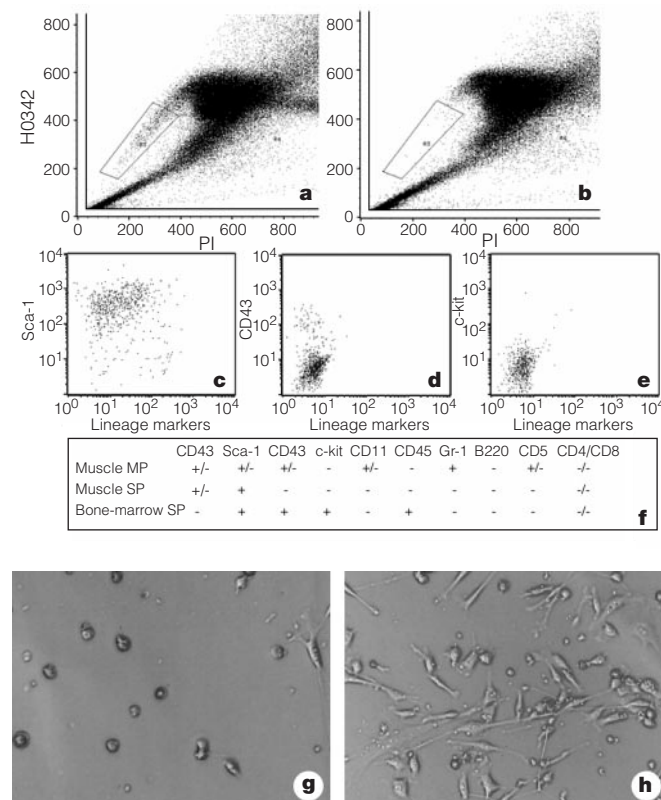


Figure 2 Isolation and characterization of muscle SP cells. A side population (SP) of cells was identified by FACS analysis (a, boxed area). In the presence of verapamil, this population was not visible (b). Over 80% of muscle SP cells express the antigen Sca-1 and are negative for lineage markers (c), CD43 (d) and c-kit (e). The antigens expressed on muscle SP and MP cells compared to bone marrow SP cells are shown in f. +/-, mixture of positive and negative cells; +, cells positive for the marker; -, cells negative for the marker. Muscle SP cells (g) and MP cells (h) in culture appear morphologically different.

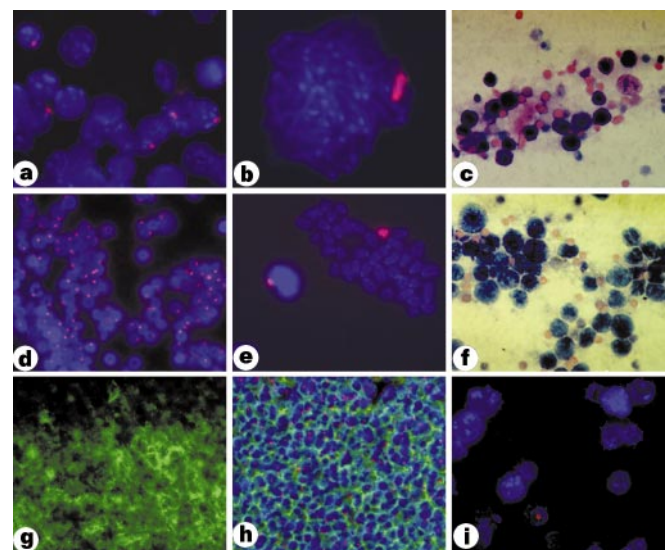


Figure 3 Detection of donor nuclei in the bone marrow and spleen of *mdx* recipients after injection of muscle SP cells. a–c, Bone marrow, animal 4; d–f, bone marrow, animal 3. Y-chromosomes are detected as red hybridization signals over the nuclei (a, d, 60× original magnification). Y-chromosome positive metaphases were detected at high magnification (b, e, 100×). Giemsa stain of bone marrow samples shows the presence of several cell types (c, f, 100×). Expression of CD43 (g) or CD45 (h) antigens in spleen tissue sections from cells of donor origin (h); animal 3. i, Detection of few male nuclei in the bone marrow of animals co-injected with muscle and bone marrow SP cells.

Table 2 Detection of donor-derived dystrophin-positive muscle fibres after transplantation of muscle SP cells

Animal	Days after injection	No. cells injected	% Dys ⁺ myofibres (total)*	% Y ⁺ nuclei in myofibres†	% Y ⁺ nuclei in bone marrow‡
1	17	7,000	9 (254)	3	41
2	21	10,000	3 (276)	7	75
3	30	20,000	6 (820)	9	91
4	30	13,000	5 (800)	6	30
5	28	19,000	ND	ND	80

* Dystrophin-positive fibres as a percentage of total number of fibres (total).

† Percentage of dystrophin-positive fibres for which a Y-chromosome-positive nucleus was detected in one given section (on average, 1 or 2 per section).

‡ Percentage of Y⁺ nuclei in the bone marrow samples detected by FISH. As a positive control for these experiments, the FISH probe was hybridized in parallel to interphase nuclei from male cells, and the hybridization efficiency was >95%.
ND: not determined.

Hoechst dye¹⁰. In contrast, the MP myoblasts were unaffected by verapamil (Fig. 2b).

Characterization of muscle SP cells revealed several unique features that distinguished them from bone marrow SP and muscle MP cells. First, isolation of muscle SP cells required a concentration of H0342 dye that was 2.5 times greater than that used to purify bone marrow SP cells¹⁰. In addition, although both muscle and bone marrow SP cells were Sca-1⁺ lin⁻, as predicted for early progenitor cells, c-Kit and CD45, two surface markers expressed on bone marrow SP cells¹¹, were not present on muscle SP cells (Fig. 2e, f). Similarly, over 90% of muscle SP cells were negative for CD43 (Fig. 2d), another marker detected on bone marrow SP cells. In contrast, muscle MP cells expressed several lineage markers (lin⁺), such as CD11, Gr-1 and CD5 (Fig. 2f). The two muscle-cell populations were also different in appearance. *In vitro* culture experiments revealed that, after one week, nearly all MP cells adhered to the culture dish and were fully differentiated into myoblasts (Fig. 2h), with a few intervening fibroblasts as determined by desmin staining (data not shown). In contrast, most SP

cells maintained a spherical shape and failed to settle on the plate (Fig. 2g). Only after 2 weeks in culture did muscle SP cells differentiate as a mixture of myoblasts and fibroblasts.

To evaluate the potential of muscle SP cells to contribute to muscle and/or haematopoietic compartments, we prepared muscle SP cells from normal C57BL/10 male mice and injected them into the tail veins of lethally irradiated female *mdx* mice. Preliminary bone marrow transplantation experiments using muscle SP cells indicated that at least ten-fold more muscle SP cells than bone marrow SP cells were needed for reliable radioprotection (data not shown). Based on those studies, 7,000–20,000 male muscle SP cells were injected into five lethally irradiated *mdx* females (Table 2). At day 17, one mouse injected with 7,000 muscle SP cells (Table 2, animal 1) appeared weak and was killed. The other animals were killed at 28–30 days, when all seemed in good health (Table 2, animals 2–5). Analysis of the bone marrow cells of the recipients by FISH indicated variable engraftment by donor cells, ranging from 30–91% (Fig. 3a, d; Table 2). Metaphase spreads containing the Y chromosome were also detected by FISH (Fig. 3b, e), indicating that the introduced male muscle SP cells could divide *in vivo*. Giemsa staining of bone marrow samples of these animals revealed diverse types of haematopoietic cell (Fig. 3c, f). In addition to bone marrow, spleen tissue sections of animals 3 and 4 were immunostained with either anti-CD43 or anti-CD45 antibodies, two surface markers that are expressed by haematopoietic cells but not by muscle SP cells (Fig. 2f). Both antibodies revealed the presence of immunoreactive cells (Fig. 3g, h). Co-detection of donor nuclei by FISH showed that over 90% of the spleen cells, including those expressing CD43 or CD45, were positive for the Y chromosome and thus of donor origin (Fig. 3h). Collectively, these findings showed that muscle SP cells can reconstitute the haematopoietic compartment of lethally irradiated recipients, albeit less effectively than bone marrow SP cells. In fact, in competitive repopulation studies of lethally irradiated *mdx* females injected with a mixture of 200 bone marrow SP cells derived from *mdx* females and 6,000 muscle SP cells derived from normal males, less than 1% of the bone marrow nuclei were positive for the Y chromosome by FISH at 4 and 8 weeks (Fig. 3i).

To evaluate the ability of muscle SP cells to differentiate into muscle, bone marrow transplant recipients engrafted with muscle SP cells were analysed by immunohistochemistry combined with FISH (Table 2; Fig. 4). Analysis of 15–30 skeletal muscle tissue sections from each of four recipients revealed donor-derived dystrophin-positive myofibres, with on average one or two Y⁺ nuclei fused to dystrophin-positive fibres in each section (Fig. 4). For animals injected with muscle SP cells, analysis of 18 photographed muscle tissue sections from three different animals revealed a pattern of nuclear localization similar to that seen in animals injected with bone marrow cells. Of a total of 28 donor male nuclei within dystrophin-positive myofibres, 12 nuclei were centrally located in the dystrophin-positive myofibres (Fig. 4b–d), 9 were peripherally located (Fig. 4f, solid arrowhead) and 7 were clearly fused to the myofibres but whether they were centrally or

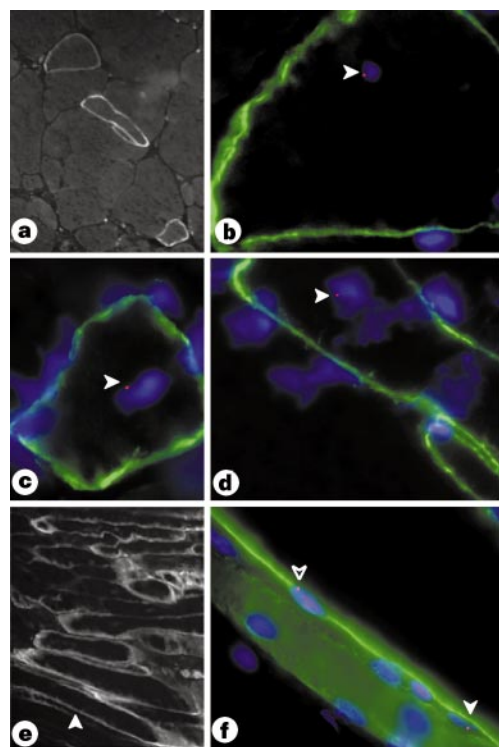


Figure 4 Dystrophin-positive fibres were detected in skeletal muscle tissue sections of animals 3 and 1. **a, e**, Low-magnification fields (original magnification (25 $\times$) showing the dystrophin-positive myofibres seen in **b–d**, respectively. **a**, Animal 3; **e**, animal 1. Solid arrowheads in **b–d, f** indicate at high magnification (100 $\times$) donor male nuclei fused to host dystrophin-positive myofibres. A donor nucleus appears to be juxtaposed to the sarcolemma of the myofibre, implying that it may be a satellite cell (**f**, open arrowhead).

peripherally located was unclear. Furthermore, a few donor nuclei were detected juxtaposed to a myofibre (Fig. 4f, open arrowhead), at a position consistent with that of satellite cells¹⁵. Analysis of serial muscle tissue sections for the expression of dystrophin and laminin showed that donor male nuclei that were juxtaposed but not fused to dystrophin-positive myofibres were encompassed by laminin staining in the adjacent sections, implying that satellite cells arise after transplantation of muscle SP cells (data not shown). In contrast, analysis of more than 200 tissue sections from animals injected with either whole bone marrow or highly purified haematopoietic cells failed to reveal the presence of Y⁺ nuclei at this position. Although these findings indicate that muscle-derived stem cells may differ from bone marrow derived stem cells in their capacity to give rise to satellite cells, more detailed studies of the precise anatomical position and molecular characteristics of these cells associated with donor-derived nuclei will be necessary to definitely address this issue.

Previous cell transplantation strategies aimed at the delivery of dystrophin have been quite successful^{4,14}, but have required multiple, local intramuscular injections, as systemic delivery of cells to muscle through myoblast transplantation does not appear to occur¹⁵. Similarly, the delivery of dystrophin to muscle by *in vivo* gene transfer with viral vectors has resulted in only local restoration of dystrophin^{16–19}. The most important practical conclusion of our studies is that bone marrow or muscle SP cells appear to provide a means for the systemic, rather than local, repair of muscle, as a consequence of the delivery of the cells throughout the vascular system. Although, in general, the proportion of dystrophin-positive myofibres that resulted from stem cell transplantation was below the levels that would be likely to be needed to provide clinical benefit in patients with muscular dystrophy, it is possible that, in the future, the procedures for stem cell transplantation that we have employed could be optimized to provide levels of engraftment of muscle that would be clinically useful^{20,21}.

Our studies also shed important new light on the biological properties of different stem cell populations. First, the experiments involving the transplantation of highly purified haematopoietic stem cells provide the first direct evidence that haematopoietic stem cells with the capacity for the complete reconstitution of lethally irradiated recipients have the potential to differentiate into muscle. Previous *in vitro* studies^{22–24} had indicated that mesenchymal stem cells of the bone marrow stroma might be the relevant source of the cells with muscle potential revealed by the earlier bone marrow transplantation studies of Ferrari *et al.* and Bittner *et al.*^{8,9}. A second important biological finding from our studies is that, using purification methods similar to those used previously to isolate haematopoietic stem cells, a population of cells from muscle can be isolated with similar functional and phenotypic properties to those of haematopoietic stem cells isolated from bone marrow. These results raise the possibility that there may be some direct relationship between bone marrow derived stem cells and other tissue- or organ-specific stem cells. One possibility that needs to be explored further is that stem cells resident in specific tissues or organs originate from a common set of cells in the bone marrow, yet adopt tissue- or organ-specific characteristics upon seeding within a specific local environment. Alternatively, organ- and tissue-specific stem cell populations may arise solely as a consequence of the normal development of that specific tissue or organ, but in general share similar phenotypic and functional characteristics. In this regard, it is interesting that we have found that SP-like cells can be isolated from a number of other differentiated tissues and organs (unpublished results). Several studies have also shown that certain bone marrow derived cells have the potential to differentiate into other non-haematopoietic cells, such as endothelial cells^{25,26}, and that some CNS-derived stem cell lines have the potential to differentiate into haematopoietic cells²⁷. Clearly, more studies to assess the origin and functional properties of different stem cell

populations and their relationship to bone marrow derived stem cells are urgently needed. □

Methods

Mouse strains

C57BL/10 mice and C57BL/10ScSn-Dmd^{mdx}/J (X-linked muscular dystrophy) mice were obtained from the Jackson Laboratory.

Purification of bone marrow and haematopoietic stem cells

Bone marrow was extracted from the femurs and tibias of 6–8-week-old animals and red blood cells were removed with ammonium chloride solution (Sigma). To purify haematopoietic stem cells, bone marrow cells were resuspended at 10⁶ cells per ml and stained with 5 µg ml⁻¹ Hoechst 33342 (Sigma) as described¹⁰. Cells were then magnetically pre-enriched for Sca-1-positive cells using the MACS (Miltenyi Biotec) and streptavidin microbeads. Before cell sorting, cells were resuspended in HBSS containing 2% FCS and 2 µg ml⁻¹ propidium iodide (PI). Flow cytometric analysis and cell sorting were performed on a dual-laser FACS Vantage flow cytometer (Becton Dickinson). Both the Hoechst and propidium iodide dyes were excited at 350 nm and their fluorescence was measured at 450 and 600 nm. The sorting gate for the haematopoietic stem cell population was established as described¹⁰.

Isolation of muscle SP cells

We isolated skeletal muscle myoblasts from 3–5-week-old donor male tissue as described¹². Before H0342 staining, red cells were lysed²⁸. Primary myoblasts were resuspended at 10⁶ cells per ml and stained with 12.5 µg ml⁻¹ of H0342 in PBS-0.5%BSA for 90 min at 37 °C. In parallel, as a negative control for SP cells, we stained 10⁶ cells in the presence of 50 µM verapamil and used them to set the gate for isolation of SP cells by FACS in the test sample¹⁰. Cells were washed once in cold PBS-0.5%BSA, resuspended at 10⁸ cells per ml and incubated for 10 min on ice with 10 µg ml⁻¹ of biotinylated anti-Sca-1 antibody (Pharmingen). For cell-lineage marker analysis, muscle SP and MP cells were stained as described¹⁰. Before FACS analysis and sorting, samples were enriched for Sca-1⁺ cells and stained with 2 µg ml⁻¹ PI¹⁰. Before being injected into animals, muscle SP cells were washed once in PBS-0.5%BSA and resuspended in 200 µl of PBS-0.5%BSA. For cell culture, muscle SP and MP myoblasts were maintained as described¹².

Bone marrow, haematopoietic or muscle stem cell transplantation

Female *mdx* recipients were lethally irradiated with 1200 rad given in two doses (600 rad each), 3 h apart. Bone marrow, haematopoietic or muscle stem cells were given intravenously into the tail vein. Mice were maintained on acidified water after transplantation. All animal care was in accordance with institutional guidelines.

Tissue collection and FISH analysis

Recipient animals were killed and skeletal muscle and spleen were snap-frozen in cold isopentane and stored at -80 °C. The bone marrow was isolated from the hind leg bones. Cells were washed in 1 × PBS and filtered through a 70 µm filter. For Giemsa staining, bone marrow cells were spread on a glass slide, fixed in methanol for 3 min and stained according to the manufacturer's instructions (Sigma). For FISH analysis of bone marrow nuclei, cells were treated with hypotonic solution and fixed in methanol and acetic acid before slide preparation as described²⁹.

We prepared the Y-chromosome FISH probe (a gift from E. Snyder) by labelling 1 µg of plasmid DNA with digoxigenin-11-dUTP as described^{29,30}. FISH was standardized on whole nuclei isolated from a male murine muscle cell line and on male muscle tissue sections. The hybridization efficiency was >90% on whole nuclei and 70–80% on tissue sections.

Immunohistochemistry and *in situ* hybridization were performed on the same tissue sections as described³⁰. Slides were examined using a Zeiss Axiophot microscope and images were collected using a CCD camera (Photometrics) as described³⁰.

Received 10 May; accepted 19 July 1999.

- Mulligan, R. C. The basic science of gene therapy. *Science* **260**, 926–932 (1993).
- Sicinski, P. *et al.* The molecular basis of muscular dystrophy in the *mdx* mouse: a point mutation. *Science* **244**, 1578–1580 (1989).
- Hall, Z. W. & Ralston, E. A. Nuclear domains in muscle cells. *Cell* **57**, 771–772 (1989).
- Karpati, G. *et al.* Dystrophin is expressed in *mdx* skeletal muscle fibers after normal myoblast implantation. *Am. J. Pathol.* **135**, 27–32 (1989).
- Pavlat, G. K., Rich, K., Webster, S. G. & Blau, H. M. Localization of muscle gene products in nuclear domains. *Nature* **337**, 570–573 (1989).
- Gussoni, E., Blau, H. M. & Kunkel, L. M. The fate of individual myoblasts after transplantation into muscles of DMD patients. *Nature Med.* **3**, 970–977 (1997).
- Hoffman, E. P., Morgan, J. E., Watkins, S. C. & Partridge, T. A. Somatic reversion/suppression of the mouse *mdx* phenotype *in vivo*. *J. Neurol. Sci.* **99**, 9–25 (1990).
- Ferrari, G. *et al.* Muscle regeneration by bone marrow-derived myogenic progenitors. *Science* **279**, 1528–1530 (1998).
- Bittner, R. E. *et al.* Recruitment of bone marrow-derived cells by skeletal and cardiac muscle in adult dystrophic *mdx* mice. *Anat. Embryol. (Berl.)* **199**, 391–396 (1999).
- Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating *in vivo*. *J. Exp. Med.* **183**, 1797–1806 (1996).

11. Goodell, M. A. *et al.* Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nature Med.* **3**, 1337–1345 (1997).
12. Rando, T. A. & Blau, H. M. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J. Cell Biol.* **125**, 1275–1287 (1994).
13. Mauro, A. Satellite cells of skeletal muscle. *J. Biophys. Biochem. Cytol.* **9**, 493–495 (1961).
14. Partridge, T. A., Morgan, J. E., Coulton, G. R., Hoffman, E. P. & Kunkel, L. M. Conversion of mdx myofibers from dystrophin negative to positive by injection of normal myoblasts. *Nature* **337**, 176–179 (1989).
15. Partridge, T. A. Invited review: myoblast transfer: a possible therapy for inherited myopathies? *Muscle Nerve* **14**, 197–212 (1991).
16. Acsadi, G. *et al.* Human dystrophin expression in mdx mice after intramuscular injections of DNA constructs. *Nature* **352**, 815–818 (1991).
17. Acsadi, G. *et al.* Human dystrophin expression in mdx mice after intramuscular injections of DNA constructs. *Nature* **352**, 815–818 (1991).
18. Ragot, T. *et al.* Efficient adenovirus-mediated transfer of a human minidystrophin gene to skeletal muscle of mdx mice. *Nature* **361**, 647–650 (1993).
19. Kochanek, S. *et al.* A new adenoviral vector: replacement of all viral coding sequences with 28 kb of DNA independently expressing both full-length dystrophin and β -galactosidase. *Proc. Natl Acad. Sci. USA* **93**, 5731–5736 (1996).
20. Rafael, J. A. *et al.* Prevention of dystrophic pathology in mdx mice by a truncated dystrophin isoform. *Hum. Mol. Genet.* **3**, 1725–1733 (1994).
21. Phelps, S. F. *et al.* Expression of full-length and truncated dystrophin mini-genes in transgenic mdx mice. *Hum. Mol. Genet.* **4**, 1251–1258 (1995).
22. Pereira, R. F. *et al.* Cultured adherent cells from marrow can serve as long-lasting precursor cells for bone, cartilage, and lung in irradiated mice. *Proc. Natl Acad. Sci. USA* **92**, 4857–4861 (1995).
23. Saito, T. *et al.* Myogenic expression of mesenchymal stem cells within myotubes of mdx mice *in vitro* and *in vivo*. *Tissue Eng.* **1**, 327–343 (1995).
24. Prockop, D. J. Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science* **276**, 71–74 (1997).
25. Asahara, T. *et al.* Isolation of putative progenitor endothelial cells for angiogenesis. *Science* **275**, 964–967 (1997).
26. Shi, Q. *et al.* Evidence for circulating bone marrow-derived endothelial cells. *Blood* **92**, 362–367 (1998).
27. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells *in vivo*. *Science* **283**, 534–537 (1999).
28. Baroffio, A., Bochaton-Piallat, M.-L., Gabbiani, G. & Bader, C. R. Heterogeneity in the progeny of single human muscle satellite cells. *Differentiation* **59**, 259–268 (1995).
29. Lichter, P. *et al.* Rapid detection of human chromosome 21 aberrations by *in situ* hybridization. *Proc. Natl Acad. Sci. USA* **85**, 9664–9668 (1988).
30. Gussoni, E. *et al.* A method to codetect introduced genes and their products in gene therapy protocols. *Nature Biotech.* **14**, 1012–1016 (1996).

Acknowledgements

We thank E. Snyder for the mouse Y-chromosome probe. This work was supported by the Muscular Dystrophy Association (L.M.K.), the Bernard and Alva Gimbel Foundation and Family (L.M.K.), the Howard Hughes Medical Institute (L.M.K., R.C.M.) and the NIH (R.C.M.). L.M.K. and R.C.M. are Investigators of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.C.M. (e-mail: mulligan@rascal.med.harvard.edu) or L.M.K. (e-mail: kunkel@rascal.med.harvard.edu).

Evidence that a free-running oscillator drives G1 events in the budding yeast cell cycle

Steven B. Haase & Steven I. Reed

Department of Molecular Biology, The Scripps Research Institute,
10550 North Torrey Pines Road, La Jolla, California 94303, USA

In yeast and somatic cells, mechanisms ensure cell-cycle events are initiated only when preceding events have been completed¹. In contrast, interruption of specific cell-cycle processes in early embryonic cells of many organisms does not affect the timing of subsequent events², indicating that cell-cycle events are triggered by a free-running cell-cycle oscillator. Here we present evidence for an independent cell-cycle oscillator in the budding yeast *Saccharomyces cerevisiae*. We observed periodic activation of events normally restricted to the G1 phase of the cell cycle, in cells lacking mitotic cyclin-dependent kinase activities that are essential for cell-cycle progression. As in embryonic cells, G1

events cycled on schedule, in the absence of S phase or mitosis, with a period similar to the cell-cycle time of wild-type cells. Oscillations of similar periodicity were observed in cells responding to mating pheromone in the absence of G1 cyclin (Cln)- and mitotic cyclin (Clb)-associated kinase activity, indicating that the oscillator may function independently of cyclin-dependent kinase dynamics. We also show that Clb-associated kinase activity is essential for ensuring dependencies by preventing the initiation of new G1 events when cell-cycle progression is delayed.

In wild-type budding yeast, completion of cell-cycle events in one cell-cycle phase is essential for the transition to the next phase¹, although mutants have been identified that can initiate synchronous, periodic rounds of budding in the absence of DNA replication and mitosis³. Strains bearing temperature-sensitive mutations in genes involved in protein degradation, *cdc4*, *cdc34* or *cdc53*, cannot enter S phase or mitosis but do initiate synchronous rounds of budding at the restrictive temperature⁴. The molecular basis for the inability of these mutants to replicate DNA or complete mitosis is related to the stabilization of the Clb/Cdc28-specific inhibitor, Sic1p, and the consequent loss of Clb-kinase activity⁵. That *cdc4*, *cdc34* and *cdc53* cells produce multiple buds indicates that the initiation of G1 events may be uncoupled from the completion of S phase or mitosis by the loss of Clb-associated kinase activity.

To investigate this idea, we analysed events normally restricted to G1 in synchronous populations of cells bearing the temperature-sensitive mutation *cdc4-3*, or cells disrupted for all six B-type cyclin genes. Synchronous populations of G1 cells were collected by centrifugal elutriation, and then shifted to the restrictive temperature (36°C). Budding and transcript levels were analysed at 10-min intervals. Synchronous oscillations of *CLN2* transcript levels were reproducibly observed with periods coinciding with the budding cycles (Fig. 1d–f, g–i). The cycling of G1 events indicates the existence of an oscillator that cycles independently of B-type cyclin activity and the completion of S phase or mitosis. The period of budding and transcription cycles in cells lacking B-type cyclin activity is very similar to the period observed in normally dividing wild-type cells (Fig. 1a–c), indicating that G1 events in normally dividing cells may be entrained to the same independent oscillator.

Activation of Cln/Cdc28 kinase is a critical G1 event that is required for the initiation of budding and the transition from G1 to S phase⁶. Therefore, we investigated whether the oscillations in *CLN2* transcript levels in cells lacking Clb activity led to functional oscillations in Cln2-associated kinase. Indeed, Cln2-associated kinase did oscillate synchronously with the appearance of new buds in *cdc4-3* cells (Fig. 1j, k). However, it is likely that the oscillation of Cln2-associated kinase activity is not essential for driving budding cycles, as constitutive expression of Cln2 from the *GAL1* promoter did not prevent rebudding in *cdc4-3* cells (data not shown). Other events essential for rebudding cycles may be entrained to the oscillator independently of Cln2-associated kinase activity (see below).

As Clns can auto-activate their own transcription^{7,8} and also appear to stimulate their own proteolysis^{9,10}, they could conceivably define an oscillator that drives G1 events in the absence of cell-cycle progression. Our finding that *cdc4-3 cdc28-4* double mutant cells do not exhibit budding or G1-specific transcription cycles indicates that Cln/Cdc28 kinase activity may be essential for these events (data not shown). However, our experiments do not distinguish whether Clns are components of an oscillator or are simply entrained to a Cln-independent oscillator, as the outputs we measured (budding and transcription) are themselves dependent on Cln/Cdc28-kinase activity.

Previous observations indicate that measurable oscillations may occur independent of Cln/Clb-associated kinase activity. In response to mating pheromone, budding yeast adopt a unique morphology, termed 'shmoo', by polarizing growth to form a

mating projection. Cells that do not conjugate with a partner of the opposite mating type can initiate multiple projections over time¹¹. The formation of mating projections is similar to budding, requiring many of the same genes¹² and involving the polarization of cortical F-actin to the projection site¹³. However, projection formation does not require Cdc28-kinase activity. To determine whether mating-projection formation occurs with a periodicity characteristic of budding cycles, we examined the formation of projections in a synchronous population of G1 cells deleted for all three *cln* genes.

We treated an elutriated G1 population of *cln1,2,3* mutant cells with the yeast mating pheromone, α -factor, and found that cells initiated multiple synchronous rounds of projection formation (Fig. 2a, c). The appearance of the second projection depends on the continued presence of α -factor (data not shown), indicating that the second projection may be a normal response to mating pheromone. As Cln activity is required for Clb expression¹⁴, *cln1,2,3* cells lack both Cln- and Clb-associated kinase activity. Thus the synchronous rounds of mating projection formation are triggered independently of Cln- and Clb-associated Cdc28-kinase activity. The kinetics of mating projection formation mirrored the kinetics of multiple bud formation in synchronized *clb1,2,3,4,5,6* mutant cells (Fig. 2a–c). Additionally, analysis of actin polarization by rhodamine–phalloidin staining showed that polarized actin is re-localized in cells immediately preceding the appearance of both new buds and new mating projections (Fig. 2b, c). Collectively, these data indicate that the relocalization of polarized growth machinery during mating projection cycles or budding cycles is entrained to the activity of a Cln/Clb-independent oscillator.

It is likely that the extended period of oscillations in these

experiments (~100 min) relative to those presented in Fig. 1 (70–80 min) reflects the lower temperature (30 °C) at which these experiments were performed. The period of budding cycles and mating-projecting cycles is similar to the cell-cycle time of wild-type cells growing at 30 °C (ref. 19), indicating that the same Cln/Clb-independent oscillator may also direct the polarized growth machinery as well as the initiation of other G1 events in normally dividing cells.

That G1 events may be triggered by the activity of an independent oscillator in normal cycling cells is not irreconcilable with previously established dependencies observed in the yeast cell cycle. When cell-cycle progression is perturbed, checkpoint signals delay progression into the next phase and may also directly or indirectly attenuate the activity of the oscillator. As we have shown that the loss of Clb-associated kinase activity uncouples the initiation of G1 functions from the completion of S phase or mitosis, we considered that Clb-associated kinase activity may be involved in preventing the activation of G1 events during a checkpoint arrest.

In budding yeast, cells arrest in mitosis with elevated Clb/Cdc28 kinase activity in response to incomplete DNA replication or spindle dysfunction^{16,17}. We investigated whether Clb/Cdc28-kinase activity is required to prevent G1-specific transcription and budding during spindle and DNA-replication checkpoint arrests. To eliminate Clb activity under checkpoint arrest conditions, we constructed strains that expressed a hyperstabilized allele of the Clb/Cdc28 inhibitor, *SIC1- Δ 3P*¹⁸, from the inducible *GAL1* promoter in wild-type cells and in a strain bearing a temperature-sensitive mutation in the thymidylate kinase gene¹⁹, *cdc8*.

In separate experiments, spindle integrity and DNA-replication checkpoint arrests were induced in the absence of *Sic1- Δ 3P* expression by the addition of the microtubule-destabilizing drug nocodazole (Fig. 3a–c), or by shifting *cdc8-1* cells to the restrictive temperature (Fig. 3d–f), respectively. When ~90% of the cells arrested with large buds, *Sic1- Δ 3P* expression was induced at 0 min

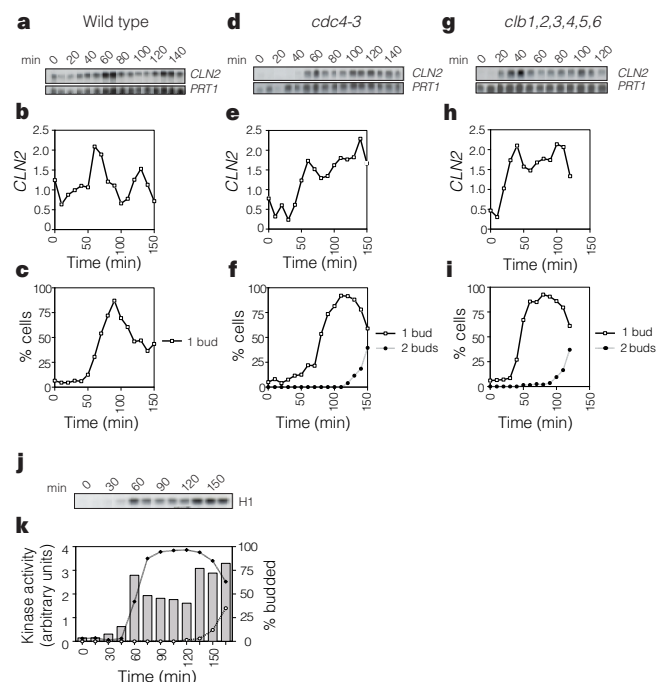


Figure 1 G1-specific transcription, budding cycles and Cln2-associated kinase activity in wild-type cells and cells lacking Clb activity. **a–i**, Time course following elutriation of G1 cells for wild-type cells (**a, c**), *cdc4-3* (**d–f**) and *clb1,2,3,4,5,6* cells (**g–i**). Analyses of *CLN2* transcripts (**a, b, d, e, g, h**) and budding indices (**c, f, i**) are shown. As the normalization transcript *PRT1* exhibits weak oscillations with the same periodicity as *CLN2* (ref. 30), the oscillation of *CLN2* is likely to be underestimated. Wild-type cells divided, so only single buds were scored. **j, k**, Synchronous oscillations of Cln2-associated kinase activity in *cdc4-3 CLN2-HA3* cells. Time course following elutriation is shown. **j**, H1-kinase activity of anti-HA precipitates. **k**, Quantification of H1-kinase activity (solid bars) and budding index (filled symbols, 1 bud; open symbols, 2 buds). All results shown are representative of multiple experiments.

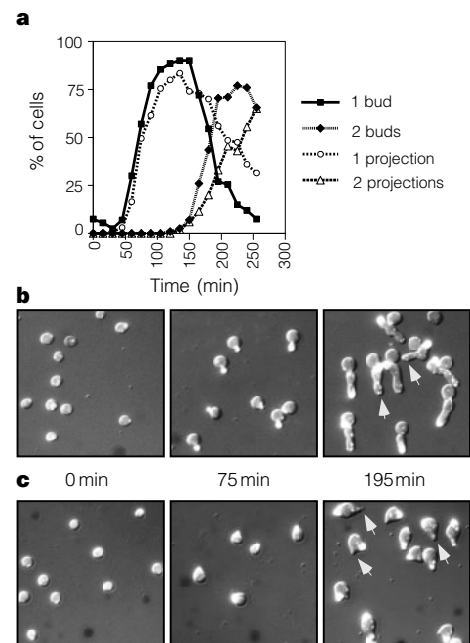


Figure 2 Synchronous mating projection cycles in a *cln1,2,3*-null mutant cells have similar kinetics to budding cycles in *clb1,2,3,4,5,6* mutant cells. Time course of synchronous G1 cells collected by elutriation. **a–c**, Mating projections were scored for *cln1,2,3* cells (**a**, open symbols; **c**) and buds were scored for *clb1,2,3,4,5,6* cells (**a**, filled symbols; **b**). Arrows indicate some sites of initial buds or mating projections (**b** and **c**, respectively). Phalloidin staining indicates presence of cortical F-actin.

while maintaining the checkpoint-arrest conditions. Upon induction of Sic1- Δ 3P, new buds emerged from the large budded cells in both the spindle and DNA-replication checkpoint-arrested cells expressing Sic1- Δ 3P, but not in control cells lacking Sic1- Δ 3P (Fig. 3c, f). Also, *CLN2* transcription was activated coincidentally with the appearance of new buds in cells expressing Sic1- Δ 3P, but not in control cells (Fig. 3a, b, d, e). DNA staining confirmed that rebudded cells had a single nucleus, indicating that mitosis had not occurred in response to the shift to galactose medium (data not shown). Similar results have been observed for cells in which Clb-associated kinase activity was eliminated in cells arrested during a spindle-checkpoint arrest^{20,21}.

These results indicate a role for B-type cyclin-associated kinase in preventing the initiation of new G1 events during checkpoint arrests, possibly by downregulating *CLN2* transcription²² or by directly attenuating the activity of the independent oscillator.

Our data support a model for the budding yeast cell cycle in which G1 events are triggered by a Cln/Clb-independent cell-cycle oscillator in normal cycling cells. That an independent oscillator may regulate some aspects of the yeast cell cycle indicates that cell-cycle regulation in embryonic cells and somatic cells may be fundamentally similar except that checkpoint mechanisms, absent in embryonic cells, exert an additional level of control in somatic cells. Our data also indicate that, in yeast, the activity of B-type cyclin-associated kinase attenuates the activity or outputs of the independent oscillator when cell-cycle progression is blocked and checkpoints are activated.

Previous findings have indicated that a *cdc2*-independent biochemical oscillator may control CO₂ production in the yeast *Schizosaccharomyces pombe*²³. The mechanism for generating periodic oscillations in yeast remains to be determined. Other known *Saccharomyces cerevisiae* cyclin/cdk systems could theoretically define an oscillator. Three members of the *PCL* family of cyclin-like genes can activate the cyclin-dependent kinase Pho85 and exhibit cell-cycle-regulated gene expression. However, as the expression of these genes is downregulated in response to α -factor²⁴, it is unlikely that they define the oscillator described in Fig. 2. Simple single-enzyme biochemical systems can generate sustained oscillations²⁵. The free-running oscillator described here, therefore, could reflect the self-generated fluctuations of an enzyme. Alter-

natively, as Ca²⁺ oscillations have been associated with cell-cycle functions in many systems²⁶, self-limiting ion-channel-regulated Ca²⁺ fluctuations might provide the basis for a cell-cycle regulatory oscillator. □

Methods

Strains, cell synchronization and checkpoint arrests

All strains are derivatives of BF264-15Dau, and were constructed by standard yeast methods. Yeast strains were grown in rich YEP medium (1% yeast extract, 2% peptone, 0.012% adenine, 0.012% uracil) containing 2% dextrose for all strains except for the galactose-dependent strains, *clb1,2,3,4,5,6 GAL1-CLB1* and *cln1,2,3 GAL1-CLN3*, which were grown in YEP medium with 2% galactose. For synchrony experiments, temperature-sensitive strains were grown at 25 °C before elutriation. All other strains were grown at 30 °C. For *clb1,2,3,4,5,6 GAL-CLB1* and *cln1,2,3 GAL1-CLN3* cells, dextrose was added to YEP + 2% galactose medium 45 min before elutriation to terminate *CLB1* and *CLN3* expression from the *GAL1* promoter. After elutriation, *clb1,2,3,4,5,6 GAL1-CLB1* cells (Fig. 1) were maintained in G1 for 80 min by treatment with 60 ng ml⁻¹ α -factor to ensure complete destruction of Clb1 and then released at 0 min. After elutriation all strains in Fig. 1 were grown in rich YEP + 2% dextrose + 1M sorbitol at 36 °C at a density of $\sim 10^7$ per ml. Sorbitol was added to stabilize cells with elongated buds. Elutriated cells in Fig. 2 were grown in YEP dextrose medium without sorbitol at 30 °C. After elutriation, *cln1,2,3 GAL1-CLN3* were treated with 60 ng ml⁻¹ α -factor.

For arrest with nocodazole, cells were pre-synchronized in S phase by being grown for 2 h in YEP + 4% raffinose medium containing 5 mg ml⁻¹ sulphamylamide and 200 μ g ml⁻¹ amethopterin (Sigma). At this time most cells were arrested in S phase with large buds. Cells were collected by centrifugation and released into YEP + 2% raffinose medium containing 15 μ g ml⁻¹ nocodazole for 2 h. *cdc8* strains were arrested in YEP + 2% sucrose medium for 2 h.

RNA analysis, kinase assays and quantification

We used standard yeast RNA methods²⁷. Kinase assays were performed on immunoprecipitates at 25 °C, as described¹⁵. Hybridizations and H1 kinase assays were quantified using a Phosphorimager (Molecular Dynamics). Transcript signals were normalized to the levels of the constitutive *PRT1* transcript²⁸ as indicated.

Microscopy

Cells were stained for actin using rhodamine-conjugated phalloidin as described²⁹. Differential interference (DIC) and fluorescence images were collected using a Nikon Eclipse E800 microscope with a 60 $\times$ objective, and a Quantix CCD camera (Photometrics).

Received 26 May; accepted 15 July 1999.

- Hartwell, L. H. & Weinert, T. A. Checkpoints: Controls that ensure the order of cell cycle events. *Science* **246**, 629–634 (1989).
- Murray, A. W. & Kirschner, M. W. Dominoes and clocks: the union of two views of the cell cycle. *Science* **246**, 614–621 (1989).
- Hartwell, L. H. Genetic control of the cell division cycle in yeast. *Exp. Cell Res.* **69**, 265–276 (1971).
- Mathias, N. et al. Cdc53p acts in concert with Cdc4p and Cdc34p to control the G1-to-S-phase transition and identifies a conserved family of proteins. *Mol. Cell. Biol.* **16**, 6634–6643 (1996).
- Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40^{SIC1} controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**, 233–244 (1994).
- Reed, S., Wittenberg, C., Lew, D., Dulic, V. & Henze, M. G1 control in yeast and animal cells. *Cold Spring Harb. Symp. quant. Biol.* **56**, 61–67 (1992).
- Dirick, L. & Nasmyth, K. Positive feedback in the activation of G1 cyclins in yeast. *Nature* **351**, 754–757 (1991).
- Cross, F. R. & Tinklenberg, A. H. A potential positive feedback loop controlling CLN1 and CLN2 gene expression at the start of the yeast cell cycle. *Cell* **65**, 875–883 (1991).
- Tyres, M., Tokiwa, G., Nash, R. & Futcher, B. The Cln3-Cdc28 kinase complex of *S. cerevisiae* is regulated by proteolysis and phosphorylation. *EMBO J.* **11**, 1773–1784 (1992).
- Lanker, S., Valdivieso, M. H. & Wittenberg, C. Rapid degradation of the G1 cyclin Cln2 induced by CDK-dependent phosphorylation. *Science* **271**, 1597–1601 (1996).
- Buckling-Throm, E., Duntze, W., Hartwell, L. H. & Manney, T. R. Reversible arrest of haploid yeast cells in the initiation of DNA synthesis by a diffusible sex factor. *Exp. Cell Res.* **76**, 99–100 (1973).
- Chenevert, J., Valtz, N. & Herskowitz, I. Identification of genes required for normal pheromone-induced cell polarization in *Saccharomyces cerevisiae*. *Genetics* **136**, 1287–1296 (1994).
- Gehring, S. & Snyder, M. The *SPA2* gene of *Saccharomyces cerevisiae* is important for pheromone-induced morphogenesis and efficient mating. *J. Cell Biol.* **111**, 1451–1464 (1990).
- Surana, U. et al. The role of CDC28 and cyclins during mitosis in the budding yeast *S. cerevisiae*. *Cell* **65**, 145–161 (1991).
- Basco, R. D., Segal, M. D. & Reed, S. I. Negative regulation of G1 and G2 by S-phase cyclins of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 5030–5042 (1995).
- Stueland, C. S., Lew, D. J. & Reed, S. I. Full activation of p34^{CDC28} histone H1 kinase activity is unable to promote entry into mitosis in checkpoint-arrested cells of the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 3744–3755 (1993).
- Sorger, P. K. & Murray, A. W. S-phase feedback control in budding yeast independent of tyrosine phosphorylation of p34^{CDC28}. *Nature* **355**, 365–368 (1992).
- Verma, R. et al. Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Scalfani, R. A. & Fangman, W. L. Yeast gene *CDC8* encodes thymidylate kinase and is complemented by the herpes thymidine kinase gene *TK*. *Proc. Natl Acad. Sci. USA* **81**, 5821–5825 (1984).
- Schwab, M., Lutum, A. S. & Seufert, W. Yeast Hct1 is a regulator of Clb2 cyclin proteolysis. *Cell* **90**,

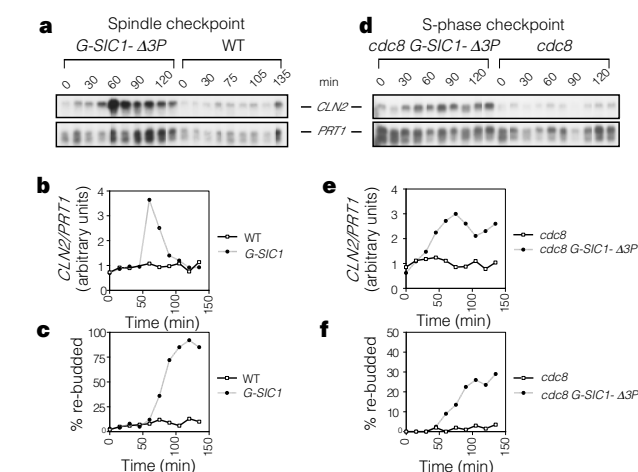


Figure 3 Clb activity is required to prevent the activation of G1 transcription. **a–c**, Time course of *GAL-SIC1-Δ3P* and wild-type (WT) cells arrested at the spindle checkpoint by nocodazole treatment. Cells were shifted to galactose medium containing nocodazole at $t = 0$ min to induce Sic1- Δ 3P expression. **d–f**, Time course of *cdc8 GAL-SIC1-Δ3P* and *cdc8* control cells arrested at the DNA-replication checkpoint by shifting to restrictive temperature (36 °C). Cells were shifted to galactose medium at $t = 0$ min to induce Sic1- Δ 3P expression and maintained at 36 °C. RNA analyses (**a**, **b**, **d**, **e**) and budding indices (**c**, **f**) are shown. Large budded cells with new buds emerging were scored as rebudded.

- 683–693 (1997).
21. Dahmann, C., Diffley, J. F. & Nasmyth, K. A. S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. *Curr. Biol.* **5**, 1257–1269 (1995).
 22. Amon, A., Tyers, M., Futcher, B. & Nasmyth, K. Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell* **74**, 993–1007 (1993).
 23. Novak, B. & Mitchison, J. M. Change in the rate of CO₂ production in synchronous cultures of the fission yeast *Schizosaccharomyces pombe*: a periodic cell cycle event that persists after the DNA-division cycle has been blocked. *J. Cell. Sci.* **86**, 191–206 (1986).
 24. Measday, V. *et al.* A family of cyclin-like proteins that interact with the Pho85 cyclin-dependent kinase. *Mol. Cell. Biol.* **17**, 1212–1223 (1997).
 25. Ngo, L. G. & Roussel, M. R. A new class of biochemical oscillator models based on competitive binding. *Eur. J. Biochem.* **245**, 182–190 (1997).
 26. Whitaker, M. & Patel, R. Calcium and cell cycle control. *Development* **108**, 525–542 (1990).
 27. Collart, M. A. & Oliviero, S. in *Current Protocols in Molecular Biology* Vol. 2 13.12 (Current Protocols, John Wiley and Sons, New York, 1993).
 28. Xu, H., Kim, U. J., Schuster, T. & Grunstein, M. Identification of a new set of cell cycle-regulatory genes that regulate S-phase transcription of histone genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 5249–5259 (1992).
 29. Adams, A. E. M. & Pringle, J. R. Staining of actin with fluorochrome-conjugated phalloidin. *Methods Enzymol.* **194**, 729–731 (1991).
 30. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).

Acknowledgements

We thank R. Deshaies for the stabilized Sic1-Δ3P construct; D. Stuart for the triple *cln* null mutant strain; M. Grunstein for the *HTA1/PRT1* probe; C. Wittenberg, N. Rhind and K. Sato for critical review of the manuscript; and members of the Reed laboratory for helpful discussions. This work was supported in part by the Leukemia Society of America and the NIH.

Correspondence and requests for materials should be addressed to S.I.R. (e-mail: sreed@scripps.edu).

Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination

Roger D. Johnson*, Nan Liu† & Maria Jasin*

* Cell Biology Program, Memorial Sloan-Kettering Cancer Center, and Cornell University Graduate School of Medical Sciences, 1275 York Avenue, New York, New York 10021, USA

† Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94551, USA

The repair of DNA double-strand breaks is essential for cells to maintain their genomic integrity. Two major mechanisms are responsible for repairing these breaks in mammalian cells, non-homologous end-joining (NHEJ) and homologous recombination (HR)^{1,2}: the importance of the former in mammalian cells is well established³, whereas the role of the latter is just emerging. Homologous recombination is presumably promoted by an evolutionarily conserved group of genes termed the *Rad52* epistasis group^{4–11}. An essential component of the HR pathway is the strand-exchange protein, known as RecA in bacteria⁸ or Rad51 in yeast⁶. Several mammalian genes have been implicated in repair by homologous recombination on the basis of their sequence homology to yeast Rad51 (ref. 11): one of these is human *XRCC2* (refs 12, 13). Here we show that *XRCC2* is essential for the efficient repair of DNA double-strand breaks by homologous recombination between sister chromatids. We find that hamster cells deficient in *XRCC2* show more than a 100-fold decrease in HR induced by double-strand breaks compared with the parental cell line. This defect is corrected to almost wild-type levels by transient transfection with a plasmid expressing *XRCC2*. The repair defect in *XRCC2* mutant cells appears to be restricted to

recombinational repair because NHEJ is normal. We conclude that *XRCC2* is involved in the repair of DNA double-strand breaks by homologous recombination.

Similar to yeast mutants that affect DNA double-strand break (DSB) repair by HR⁶, hamster cells that lack *XRCC2* are hypersensitive to ionizing radiation (about 2-fold) and crosslinking agents (60- to 100-fold), and show an increase in chromosome instability^{13,14}. In contrast to yeast, all characterized mammalian DSB-repair mutants have been found to be defective in NHEJ. Thus, the role of *XRCC2* in DNA repair is unclear. To determine whether the hamster cell line *irs1*, which is deficient in *XRCC2* (refs 12, 13), can repair DSBs by HR, we used a novel recombination reporter substrate SCneo (Fig. 1a). SCneo contains two nonfunctional copies of the neomycin phosphotransferase (*neo*) gene. One copy, designated 3' *neo*, is a 5' truncation of the *neo* gene¹⁵. The second copy, designated S2neo, is mutated at an *NcoI* site by deletion of 4 base pairs (bp) of *neo* gene coding sequence and insertion of the 18-bp site for the rare-cutting *I-SceI* endonuclease¹⁶. The two *neo* genes are in direct orientation and are separated by a functional hygromycin

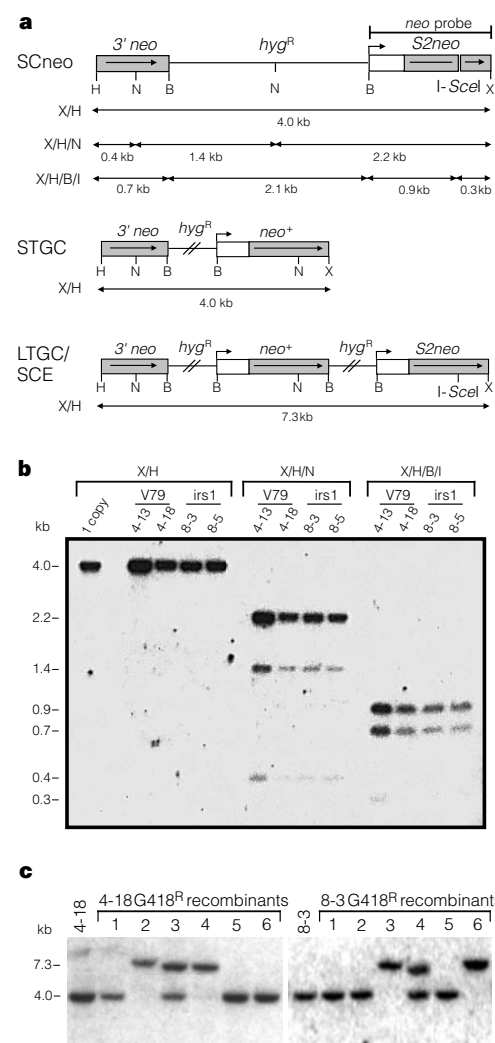


Figure 1 Recombination reporter substrate SCneo. **a**, Structure of SCneo and predicted HR products. The *neo* probe is indicated. X/H, *XhoI/HindIII*; X/H/N, *XhoI/HindIII/NcoI*; X/H/B/I, *XhoI/HindIII/BamHI/I-SceI*. **b**, Southern blot analysis of SCneo cell lines. Each cell line contains a single copy of SCneo, except the parental cell line 4-13 which contains two copies. **c**, Southern blot analysis of cell lines 4-18 (V79) and 8-3 (*irs1*) and G418^R recombinants derived from them. Genomic DNA was digested with *XhoI/HindIII* to distinguish STGC (4.0 kb) and LTGC/SCE (7.3 kb). Recombinants with both fragments probably underwent two recombination events.

resistance gene (*hyg^R*). SCneo was stably integrated into the genome of the *irs1* mutant line and its radioresistant parental line V79 by selecting for *hyg^R* cells. Clones containing an intact SCneo (two each for V79 and *irs1*) were verified by Southern blotting (Fig. 1b). In each of the clones, SCneo integrated at a different genomic location (data not shown).

In cell lines containing SCneo, DSBs introduced into the chromosome by *I-SceI* could be repaired by HR to restore a *neo⁺* gene. Wild-type cell lines transfected with the *I-SceI* expression vector pCMV3xnl*s-I-SceI* underwent HR at a frequency of $1-2 \times 10^{-3}$ per plated cell, more than 100-fold higher than cells transfected with the control plasmid (Fig. 2). The actual HR frequency is even higher, as equal sister-chromatid HR events are undetectable in our assay. This large induction of recombination by a chromosomal DSB is consistent with what has previously been reported in other systems¹⁷. In stark contrast, the *XRCC2*-deficient *irs1* cell lines had a much reduced frequency of recombinational repair, in the range of $2-6 \times 10^{-6}$. The recombination defect observed in the mutant lines can be attributed to the defect in *XRCC2*, as cotransfection of pCMV3xnl*s-I-SceI* with pXR2, a human *XRCC2* expression vector, resulted in nearly wild-type recombination levels (about 5×10^{-4} ; Fig. 2). *XRCC2* expression is evidently promoting DSB-induced recombination, because transfection of pXR2 alone is not sufficient to increase recombination (data not shown). In contrast to *XRCC2*, expression of human *Rad51* did not correct the HR defect in *XRCC2* (R.D.J. and M.J., unpublished results).

SCneo allows the detection of two recombination products (Fig. 1a). One product results from a short tract gene conversion (STGC) in which the *neo* sequences immediately surrounding the DSB in the *S2neo* gene are repaired without changing the overall architecture of the SCneo reporter (Fig. 1a). In an STGC event, repair can occur from the 3' *neo* gene located on either the same chromatid or the sister chromatid. The second predicted recombination product results in expansion of the SCneo reporter from two to three *neo*-containing repeats, either as a result of a long tract gene conversion (LTGC) with the sister chromatid or a sister-chromatid exchange (SCE) event (Fig. 1a). As shown in Fig. 1c and Table 1, both products are recovered in equal proportions from the mutant and wild-type cell lines, showing that loss of *XRCC2* reduces both STGC and LTGC/SCE recombination.

Mammalian cells possess a potent end-joining activity which repairs DSBs using little or no homology³. To determine whether *XRCC2* affects all DSB-repair processes rather than HR specifically, we used a polymerase chain reaction (PCR) based assay that simultaneously detects both NHEJ and HR (Fig. 3a; ref. 2). A DSB was induced at the SCneo substrate by transfection of the *I-SceI* expression vector, and genomic DNA was isolated immediately after transfection as a control, and after 48 h. As shown in Fig. 3b, the product amplified from genomic DNA recovered immediately after transfection (0 h) was cleaved by *I-SceI*. By contrast, much of the product amplified from genomic DNA recovered 48 h after

Table 1 Summary of DSB-induced recombination products

Cell line	No. STGC events	No. LTGC or SCE events
V79		
4-13	17	8
4-18	6	16
Total	23	24
<i>irs1</i>		
8-3	16	8
8-5	4	11
Total	20	19

transfection was not cleavable by *I-SceI*, indicating recovery of DSB-repair products. To identify HR (*NcoI*⁺) and NHEJ (*NcoI*⁻/*I-SceI*⁻) repair products, the resulting PCR products were digested with *NcoI* and/or *I-SceI* (Fig. 3b). In the wild-type cell lines, the *NcoI*⁺ fragments are readily detectable in the 48 h samples, as is the *NcoI*⁺/*I-SceI*⁻ band, indicating robust homologous and nonhomologous DSB repair. In the *XRCC2*-deficient cell lines, the *NcoI*⁺ fragments are significantly reduced (Fig. 3b), although they can be restored by complementation (data not shown). This confirms that loss of *XRCC2* results in a defect in HR. Unlike the HR product, the *NcoI*⁺/*I-SceI*⁻ NHEJ products are readily detectable in the *XRCC2* mutant, indicating that loss of *XRCC2* does not affect nonhomologous repair. Thus, the repair defect in the *XRCC2*-deficient cell line can be attributed to a defect in HR, rather than a global defect in DSB repair. Wild-type V79 cells exhibit an increased resistance to radiation during S-phase, whereas *irs1* cells do not¹⁸. Taken together, our results strongly indicate that S-phase radiation resistance in wild-type cells is due to sister-chromatid recombination, with the *irs1* cells having a defect in this type of DSB repair.

Loss of *Rad51* in mice results in early embryonic lethality¹⁹, and cells recovered from the *Rad51*^{-/-} mutant embryos contain chromosomal abnormalities²⁰. This indicates that recombinational repair may be critical for chromosome integrity and cell proliferation. *XRCC2* in the *irs1* cell line is essentially a null allele¹³, and cells lacking *XRCC2* show only a mild growth defect in culture, and are not completely defective for HR (Fig. 1c); therefore, *XRCC2* is not essential for recombinational repair or cell viability. As the mammalian *Rad51* protein has the greatest identity to yeast *Rad51* and *E. coli* *RecA*¹¹, and can catalyse strand exchange *in vitro*²¹, *Rad51* is probably the central strand-transferase protein in the cell. We speculate that *XRCC2*, which has only 20% sequence identity to *Rad51*, and other *Rad51*-related proteins may promote, but are not essential for, *Rad51* activity. This may also be true of other recombination proteins (for example, *Rad54* and *Rad52*)^{5,7}; consistent with the *Rad51*-related proteins modifying *Rad51* function, these proteins interact with one another^{13,22}, similar to the interactions observed between yeast *Rad51* and the *Rad55* or *Rad57* proteins^{23,24}, which promote *Rad51* activity *in vitro*²⁵.

DNA-repair processes are essential in maintaining chromosomal structure and genetic integrity. The significance of these processes is

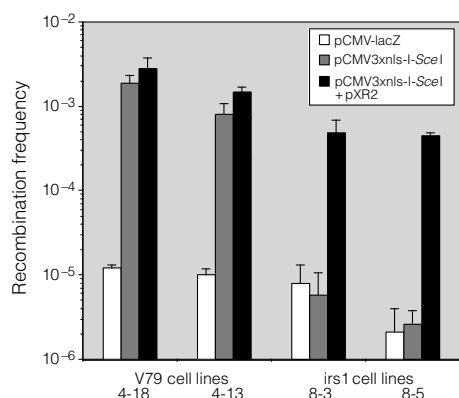


Figure 2 *XRCC2*-deficient *irs1* cells have severely reduced levels of DSB repair by HR. Transfection of wild-type V79 cell lines with the *I-SceI* expression vector increases HR more than 100-fold compared with transfection of the control plasmid pCMV-*lacZ*. By contrast, mutant *irs1* cell lines show little or no increase in HR. DSB-induced HR is substantially restored in the mutant cell lines by transient transfection of an *XRCC2* expression vector (pXR2) with the *I-SceI* expression vector. The variability observed within each pair of cell lines is probably due to position effects.

emphasized by links between defects in DNA-repair pathways and human disease and malignancy²⁶. Cancer cells frequently contain abnormal genomes exhibiting chromosomal rearrangements and aneuploidy. Thus it seems likely that failure to repair a DSB or its illegitimate repair may contribute to a cell's progression towards malignancy. The demonstration that the products of the hereditary breast cancer genes interact with Rad51 (ref. 27) indicates that recombinational repair may be disrupted in cells from these patients, and that loss of HR may promote tumorigenesis. A number of genes potentially involved in HR have been identified in mammalian cells because of their homology with yeast genes. However, the difference in severity of phenotype between yeast and mammalian cells that carry mutations in those genes (for example, Rad51 (refs 19, 20) and Rad52 (ref. 7) makes it vital to analyse recombination in mammalian cell mutants. We have demonstrated the involvement of a mammalian repair protein in recombinational repair of DNA damage; it will be important to determine whether XRCC2 and other genes involved in recombinational repair are tumour suppressor genes, as are the hereditary breast cancer genes. □

Methods

Plasmids

SCneo was constructed from IRneo, which is identical to DRneo¹⁵ except that the 3' neo gene is in reverse orientation (M.J., unpublished results). IRneo was modified by the insertion of a double-stranded oligonucleotide containing *Cla*I and *Bam*HI sites into its *Xho*I site. This oligonucleotide regenerated a single *Xho*I site and inserted *Cla*I and *Bam*HI sites upstream of the promoter of *S2neo*. This plasmid was digested with *Cla*I, which contains the *S2neo* gene on a 1.1-kb *Cla*I fragment, and then re-ligated with the *S2neo* fragment to obtain a plasmid with *S2neo* in the opposite orientation to that present in IRneo, designated SCneo. The I-SceI expression vector, pCMV3xns-I-SceI, is a modified version of pCMV-I-SceI (ref. 28) that contains a triplicated nuclear localization signal fused to I-SceI (ref. 29). Expression of I-SceI is not toxic to cultured mammalian cells²⁸. pXR2 was constructed by digesting pCDNA-XR2 (ref. 13) with *Sma*I and *Bst*BI to remove a 900-bp fragment containing the entire *neo* coding sequence. XRCC2 expression does not alter the transient expression of cotransfected genes, as *lacZ* expression is not affected when pCMV-*lacZ* is cotransfected with pXR2 (data not shown).

DNA manipulations

Southern blot analysis was performed using 8 µg genomic DNA according to standard procedures. The probe was a 1.1-kb *Xho*I–*Hind* III fragment containing the complete *neo*

gene¹⁶. PCR was done as described², using genomic DNA that has been precleaved with I-SceI *in vitro* to reduce the amplification of I-SceI⁺ *neo* genes.

Cell transfections

Transfections were done by mixing 1.6×10^7 cells suspended in PBS with 30 µg uncut plasmid and electroporating at 250V, 960 µF. Drug selections were begun 24 h after transfection. To establish cell lines with a stably integrated SCneo reporter substrate, cells were electroporated with the SCneo plasmid and selected in media containing 0.5 mg ml⁻¹ hygromycin. *hyg*^R colonies were screened by Southern blot analysis for intact incorporation of the SCneo reporter substrate. Recombination frequencies were determined by transfection of SCneo-containing cell lines with the indicated plasmid, followed by selection in media containing 1 mg ml⁻¹ G418. Cells were grown for 10–12 days and surviving G418^R colonies were stained with a 10% Giemsa solution for colony counts. Of the G418^R colonies obtained in these transfections, 95% were also *hyg*^R.

Received 15 June; accepted 27 July 1999.

- Rouet, P., Smith, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
- Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 5172–5177 (1998).
- Jeggo, P. A. DNA breakage and repair. *Adv. Genet.* **38**, 185–218 (1998).
- Bezzubova, O., Silbergleit, A., Yamaguchi-Iwai, Y., Takeda, S. & Buerstedde, J. M. Reduced X-ray resistance and homologous recombination frequencies in a RAD54–/– mutant of the chicken DT40 cell line. *Cell* **89**, 185–193 (1997).
- Essers, J. *et al.* Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. *Cell* **89**, 195–204 (1997).
- Nickoloff, J. A. & Hoekstra, M. F. in *DNA Damage and Repair Vol. I* (eds Nickoloff, J. A. & Hoekstra, M. F.) 335–362 (Humana, Totowa, 1998).
- Rijkers, T. *et al.* Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. *Mol. Cell. Biol.* **18**, 6423–6429 (1998).
- Smith, G. R. in *DNA Damage and Repair Vol. I* (eds Nickoloff, J. A. & Hoekstra, M. F.) 135–162 (Humana, Totowa, 1998).
- Sonoda, E. *et al.* Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **17**, 598–608 (1998).
- Yamaguchi-Iwai, Y. *et al.* Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Mol. Cell. Biol.* **18**, 6430–6435 (1998).
- Thacker, J. A. surfeit of RAD51-like genes? *Trends Genet.* **15**, 166–168 (1999).
- Cartwright, R., Tambini, C. E., Simpson, P. J. & Thacker, J. The XRCC2 DNA repair gene from human and mouse encodes a novel member of the recA/RAD51 family. *Nucleic Acids Res.* **26**, 3084–3089 (1998).
- Liu, N. *et al.* XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Molecular Cell* **1**, 783–793 (1998).
- Jones, N. J., Cox, R. & Thacker, J. Isolation and cross-sensitivity of X-ray-sensitive mutants of V79-4 hamster cells. *Mutat. Res.* **183**, 279–286 (1987).
- Liang, F., Romanienko, P. J., Weaver, D. T., Jeggo, P. A. & Jasin, M. Chromosomal double-strand break repair in Ku80 deficient cells. *Proc. Natl Acad. Sci. USA* **93**, 8929–8933 (1996).
- Smith, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.* **23**, 5012–5019 (1995).
- Jasin, M. Double-strand break repair and homologous recombination in mammalian cells, in *DNA Damage and Repair Vol. III* (eds Nickoloff, J. A. & Hoekstra, M. F.) (Humana, Totowa, in the press).
- Cheong, N., Wang, X., Wang, Y. & Iliakis, G. Loss of S-phase-dependent radioresistance in *irs-1* cells exposed to X-rays. *Mutat. Res.* **314**, 77–85 (1994).
- Tsuzuki, T. *et al.* Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. *Proc. Natl Acad. Sci. USA* **93**, 6236–6340 (1996).
- Lim, D.-S. & Hasty, P. A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53. *Mol. Cell. Biol.* **16**, 7133–7143 (1996).
- Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP dependent homologous pairing and strand transfer reactions *in vitro*. *Cell* **87**, 757–766 (1996).
- Dosanjh, M. K. *et al.* Isolation and characterization of RAD51C a new human member of the RAD51 family of related genes. *Nucleic Acids Res.* **26**, 1179–1184 (1998).
- Hays, S. L., Firmenich, A. A. & Berg, P. Complex formation in yeast double-strand break repair: Participation of Rad51, Rad52, Rad55, and Rad57 proteins. *Proc. Natl Acad. Sci. USA* **92**, 6925–6929 (1995).
- Johnson, R. D. & Symington, L. S. Functional differences and interactions among the putative RecA homologs Rad51, Rad55, and Rad57. *Mol. Cell. Biol.* **15**, 4843–4850 (1995).
- Sung, P. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev.* **11**, 1111–1121 (1997).
- Vogelstein, B. & Kinzler, K. W. (eds) *The Genetic Basis of Human Cancer* (MacGraw-Hill, New York, 1998).
- Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763 (1997).
- Rouet, P., Smith, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91**, 6064–6068 (1994).
- Donoho, G., Jasin, M. & Berg, P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol. Cell. Biol.* **18**, 4070–4078 (1998).

Acknowledgements

We thank L. Thompson and N. Jones for their assistance. This work was funded by an NRSA fellowship to R.D.J. a DOE grant (N.L.) and an NIH grant to M.J.

Correspondence and requests for materials should be addressed to M.J. (e-mail: m-jasin@ski.mskcc.org).

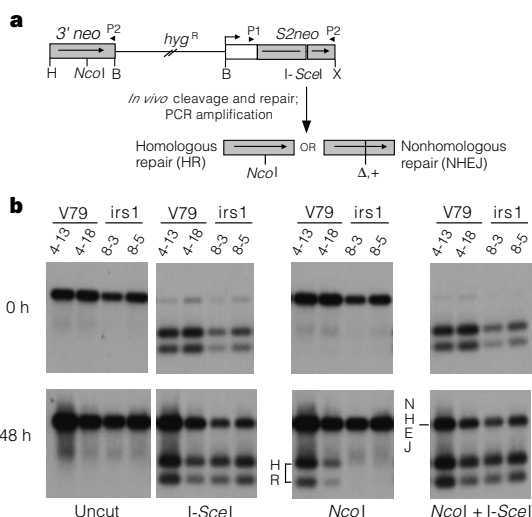


Figure 3 XRCC2-deficient *irls1* cells have normal levels of NHEJ. **a**, PCR-based assay of DSB repair. The primers amplify both products of HR occurring by either STGC or LTGC/SCE and most NHEJ events that contain small deletions and insertions (Δ , +). **b**, DSB repair in the wild-type and XRCC2-deficient cell lines. DSB-induced HR is reduced in the XRCC2 mutant, as indicated by the lack of detectable *Nco*I PCR product; however, NHEJ is robust in both the wild-type and XRCC2 mutant cell lines, as seen by the presence of the *Nco*I–I-SceI band.

Interpreting the folding kinetics of helical proteins

Yaoqi Zhou* & Martin Karplus*†

* Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA

† Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France

The detailed mechanism of protein folding is one of the major problems in structural biology^{1,2}. Its solution is of practical as well as fundamental interest because of its possible role in utilizing the many sequences becoming available from genomic analysis³. Although the Levinthal paradox⁴ (namely, that a polypeptide chain can find its unique native state in spite of the astronomical number of configurations in the denatured state) has been resolved^{4–7}, the reasons for the differences in the folding behaviour of individual proteins remain to be elucidated. Here a C_{α} -based three-helix-bundle-like protein model with a realistic thermodynamic phase diagram is used to calculate several hundred folding trajectories. By varying a single parameter, the difference between the strength of native and non-native contacts, folding is changed from a diffusion–collision mechanism⁸ to one that involves simultaneous collapse and partial secondary-structure formation, followed by reorganization to the native structure. Non-obligatory intermediates are important in the former, whereas there is an obligatory on-pathway intermediate in the latter. Our results provide a basis for understanding the range of folding behaviour that is observed in helical proteins.

All-atom models with explicit solvent have provided information on high-temperature unfolding^{9,10} and the free-energy landscapes of

several proteins^{11,12}. However, corresponding simulations of the folding kinetics for ensembles of trajectories are beyond what can be done with available computers¹³, as folding times are milliseconds or longer for most proteins. To overcome such computational limitations, we use discrete molecular dynamics to study the folding of the small (46-residue) three-helix-bundle fragment of protein A (ref. 14; and Fig. 1a). The thermodynamic behaviour of the model (Fig. 1b) is in accord with simulations of crambin using an all-atom representation and the temperature-dependent X-ray diffraction results for ribonuclease A (ref. 15), as well as data for other proteins^{16,17}. An important element of the model¹⁸ is that the phase diagram can be varied by the choice of a single parameter, the bias gap (see Methods), which determines the relative stability of native and non-native contacts. For example, equilibrium collapse to a disorganized globule occurs for a small bias gap and to an organized (molten) globule for a large bias gap, as shown in Fig. 1b. To analyse the folding kinetics, 100 trajectories, each 100 μ s to 1 ms in length, were calculated for a series of values of the bias gap parameter, starting from the fully denatured state under conditions where the native state is stable (see Methods). The folding times range from 60 ns to more than 0.1 ms for the highly optimized model (large bias gap, $g = 1.3$) and from 2 μ s to more than 1 ms for a weakly optimized model ($g = 0.3$).

The range of folding behaviour is illustrated in Fig. 2. The kinetics of two trajectories for a small value of the bias parameter ($g = 0.3$), both are very different, although identical model parameters are used and both reach the native state. The first (Fig. 2a) has early helix formation (that is, about 70% of the helical contacts exist at 10 ns), whereas other progress variables (the fraction of native interhelical contacts and the inverse native fraction of the volume, for example) are still very far from the native state. The resulting near-helical peptide chain then undergoes a slow diffusion–collision-like search⁸ of the relative helix orientations until the native state is reached at 19 μ s. In this folding scheme, there can be

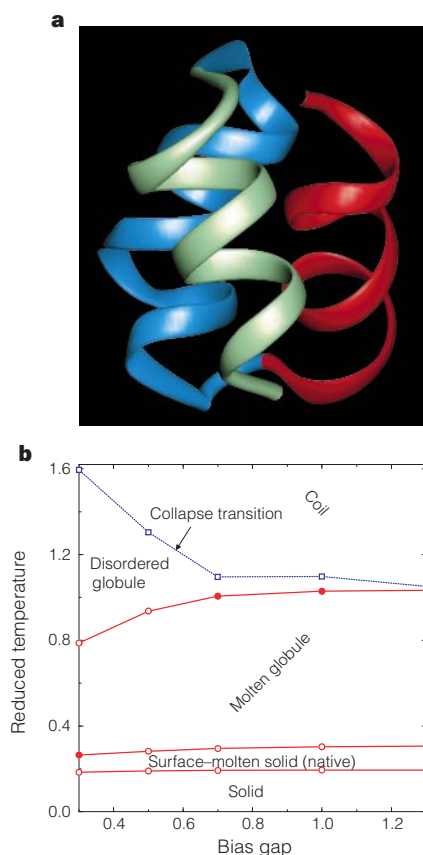


Figure 1 Structure and thermodynamics of the model three-helix bundle protein.

a, Global minimum structure of the model obtained by annealing the NMR structure of the three-helix-bundle fragment (residues 10–55) of *Staphylococcus aureus* protein A (ref. 14); helix I, II and III are shown in green, blue and red, respectively. **b**, The phase diagram for models with different bias gap values. All temperatures are in units of the energy parameter ϵ . Thermal transitions and collapse transition are indicated by open circles and open squares, respectively. The filled circles indicate a first-order-like two-state transition. There are four significant transitions: a collapse transition; a disordered to ordered (molten) globule transition; a globule to native-state transition; and the transition from the active native state to a frozen inactive state^{15,22}.

intermediates in which two helices are packed roughly as in the native state (for example, I and II), and the third (for example, III) is extended in a structure appropriate for domain-swapped dimer formation¹⁹ (structure at 1 μ s; Fig. 2a). Such a folding trajectory is rare for small bias gap values, but it dominates if the bias gap is large ($g \geq 1$). A trajectory corresponding to the predominant mechanism for the small bias gap is shown in Fig. 2b. It involves simultaneous rapid partial helix formation and collapse (90% at $t \approx 200$ ns) to a globule that has transient formation of about 20% of the native tertiary contacts, some of which have to be broken to reach the native state. The collapse is followed by a slow (~ 500 μ s) search for the native state within the compact globule; the helical content increases to its final value as the correct tertiary structure is formed.

As found in unfolding simulations of all-atom models²⁰, the folding of the three-helix-bundle model involves broad ensembles of trajectories, but they can be classified into dominant pathways (Fig. 3). The large-gap (highly optimized) model has a 'fast track' with no intermediates (folding time ~ 0.1 μ s) and slower tracks with intermediates, which require partial unfolding. The kinetic intermediates involve incorrect arrangements of the three helices (I_1) (Fig. 2a) or locally misfolded regions (I_2). In contrast, the small-gap model has an on-pathway intermediate (I'_1) with a wide range of structures, including varying amounts of helix formation (Fig. 2). I'_1 is an obligatory on-pathway intermediate because it is present in all trajectories, both folded and not folded²¹. 'Fast-track' folding is exponential, whereas the folding with intermediates is non-exponential, as expected.

The long-lived intermediates (I_1 , I'_1 and I_2) found in the simulations (Fig. 3) involve a distribution of misfolded regions that trap the system. Specifically, the C_α pseudodihedral angles for one or more residues are in the wrong region. A wide range of lifetimes for misfolded intermediates occur in the simulations; they vary from several nanoseconds to longer than the simulation time. The escape

generally involves on-site pseudodihedral angle transitions. This indicates that successful folding may be possible only at temperatures that are high enough to permit such transitions^{15,22}. The native contacts of the kinetic intermediates (I_1 , I'_1 and I_2) are in the range of the corresponding equilibrium-disordered and molten globule states, but the former have more non-native contacts.

The collapsed intermediate I'_1 shown in Fig. 2b (structures at 10 and 100 μ s) is similar to the metastable intermediate in ref. 13; both have a radius of gyration that is near the native value, and partially formed helical segments (40–60%) with approximately the correct positions and several pseudodihedral angles whose signs are opposite to the native values. The present results can be related also to all-atom calculations of the equilibrium free-energy surface for folding of the same three-helix-bundle fragment. There are two sets of populations^{11,12}: in one, helix formation is early (that is, it occurs in the state without collapse); in the other, there is little helical content in the uncollapsed state. These two dominant regions on the free-energy surface correspond to the kinetic behaviour of the weakly optimized off-lattice model.

The complete description of the folding kinetics and thermodynamics of this model system suggests some general conclusions concerning the folding mechanisms of helical proteins. In particular, a protein that is optimized for fast folding has its rate limited by the collapse transition; in other words, the intrahelical (secondary) contacts can be formed rapidly, even in the absence of collapse, and interhelical (tertiary) structural elements are formed correctly as collapse proceeds. However, misfolding and non-obligatory intermediates are likely to occur and slow down overall folding. Thus, an essential element for fast folding is elimination of non-obligatory intermediates that can act as traps so that only the fast track is present. For helical proteins that are less well optimized, the kinetic intermediates change from non-obligatory to obligatory with the appearance of the disordered globule phase (Fig. 1b). In terms of

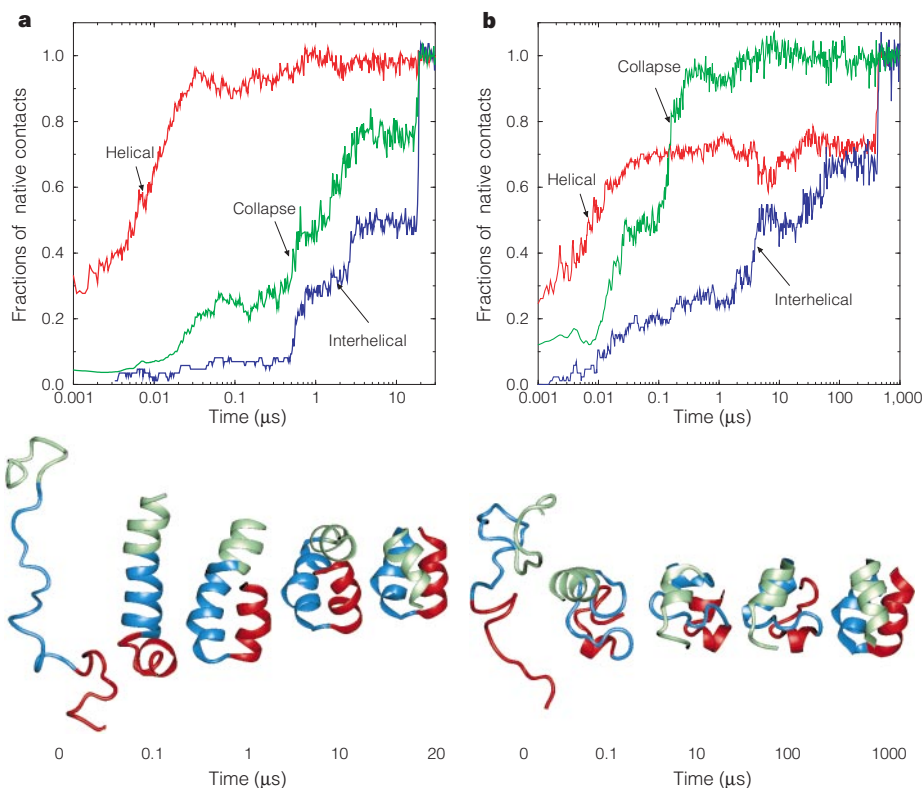


Figure 2 A semi-log plot of the time dependence of the fractions of native helical and interhelical contacts and inverse fraction of native volume (calculated from the inverse cube of the radius of gyration) for trajectories together with structures at selected times for

the $g = 0.3$ model. **a, b**, Two individual trajectories; the final structure is the native structure.

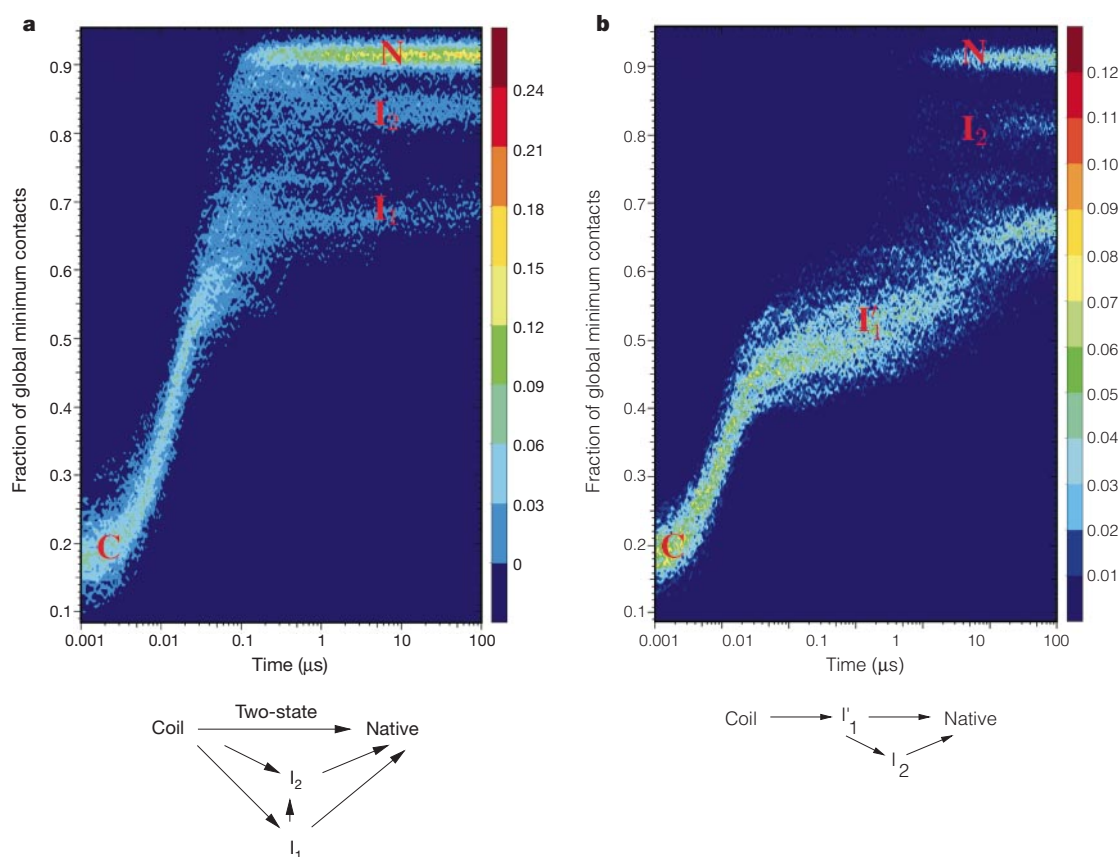


Figure 3 Folding mechanism of the large-gap ($g = 1.3$, **a**) and small-gap ($g = 0.3$, **b**) models. The distribution of the fraction of global minimum contacts is shown as a function of time (top). The distributions are obtained from independent kinetic simulations; the colour scale indicates the probabilities at given times. The various species (C, coil state; I, intermediates; N, native state) that are important in folding pathways are indicated in red.

phenomenological descriptions, the folding changes from the diffusion–collision⁸ to simultaneous collapse and secondary structure formation with more local rearrangements to the native tertiary structure.

There have been many experimental studies of the folding of α -helical proteins. The fast-folding λ -repressor (0.2 ms)²³, which is well described by the diffusion–collision model, corresponds to the fast track in the large gap limit. By contrast, folding of apomyoglobin²⁴ involves an obligatory intermediate without a precise structure and folds more slowly (0.5 s), in accord with a small gap optimization. In four-helix-bundle immunity proteins²⁵, the less stable (small gap) protein (Im7) folds more slowly in the presence of kinetic intermediates, whereas the more stable (large gap) protein (Im9) folds via faster two-state mechanism. These illustrative comparisons indicate the potential utility of the present model for understanding how folding scenarios are realized in specific α -helical proteins. It will be interesting to investigate other protein topologies (all- β , α/β) by the same approach. $\square$

Methods

Model

The model¹⁸ consists of 46 beads, each representing an amino-acid residue. The interaction potential between two non-bonded beads i and j is given by a square-well or square-shoulder potential. A Go-like model²⁶ is used with the square-well depth (or square-shoulder height) for a pair of residues equal to $B_N\epsilon$ if the pair is in contact in the global minimum structure and $B_O\epsilon$ ($B_O > B_N$), otherwise. The 'bias gap' g , $g = 1 - B_O/B_N$, is a measure of the difference in stability between the global minimum (native) contacts and other contacts. For $g > 1$, non-native contacts are repulsive ($B_O > 0$). The energy scale is determined by the parameter ϵ . The model differs from earlier work¹⁸ by the presence of a chirality bias toward right-handed helices. A potential for the pseudodihedral angle α_i

between the corresponding C_α atoms $i, i+1, i+2$ and $i+3$ is introduced. It has an energy ϵ_b ($=4\epsilon$) for $-180 < \alpha_i < 0$ and 0 for $0 < \alpha_i < 180$. The dihedral potential is applied to the α_i that are positive (right-handed) in the global minimum structure.

Thermodynamics

The thermal transition temperatures in the phase diagram are based on heat-capacity maxima, which are obtained by the weighted-histogram method²⁷ and constant-temperature discrete molecular dynamics²⁸ at 17 different temperatures; five independent equilibrium simulations with 20–200-million collision steps each were conducted at every temperature. The collapse transition temperature is determined from the temperature at which the temperature derivative of the radius gyration square is a maximum (based on a spline fit).

Kinetics

Discrete molecular dynamics were used²⁸. The 100 independent kinetic folding simulations were started with the equilibrated coil state at $T^* = 2.5$. The folding kinetics were investigated by quenching the temperature to $T^* = 0.24$. A physical time scale was obtained by setting the average collapse time ($t \approx 1,000$ reduced time units) to 1 μ s, the observed time for collapse of proteins^{29,30}.

Received 26 April; accepted 29 July 1999.

1. Fersht, A. *Enzyme Structure, Mechanism & Protein Folding* 3rd edn (W.H. Freeman, New York, 1998).
2. Dobson, D. M., Šali, A. & Karplus, M. Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.* **37**, 868–893 (1998).
3. Šali, A. 100,000 protein structures for the biologist. *Nature Struct. Biol.* **5**, 1029–1032 (1998).
4. Karplus, M. The Levinthal paradox: yesterday and today. *Fold. Des.* **2**, 569–576 (1997).
5. Karplus, M. & Šali, A. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **5**, 58–73 (1995).
6. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19 (1997).
7. Soccia, N. D., Onuchic, J. N. & Wolynes, P. G. Protein folding mechanisms and the multidimensional folding funnel. *Proteins* **32**, 136–158 (1998).
8. Karplus, M. & Weaver, D. L. Protein-folding dynamics. *Nature* **260**, 404–406 (1976).
9. Li, A. & Daggett, V. Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA* **91**, 10430–10434 (1994).

10. Caflisch, A. & Karplus, M. Acid and thermal denaturation of Barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252**, 672–708 (1995).
11. Boczek, E. M. & Brooks, C. L. III. First principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**, 393–396 (1995).
12. Guo, Z. Y., Brooks, C. L. III. & Boczek, E. M. Exploring the folding free energy surface of a three-helix bundle protein. *Proc. Natl Acad. Sci. USA* **94**, 10161–10166 (1997).
13. Duan, Y. & Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 707–744 (1998).
14. Gouda, H. *et al.* Three-dimensional solution structure of the B domain of Staphylococcal protein A: Comparisons of the solution and crystal structures. *Biochemistry* **31**, 9665–9672 (1992).
15. Zhou, Y., Vitkup, D. & Karplus, M. Native proteins are surface-molten solids: Application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* **285**, 1371–1375 (1999).
16. Ptitsyn, O. B. Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83–230 (1995).
17. Privalov, P. L. Stability of proteins: Small globular proteins. *Adv. Protein Chem.* **33**, 167–241 (1979).
18. Zhou, Y. & Karplus, M. Folding thermodynamics of a model three-helix bundle protein. *Proc. Natl Acad. Sci. USA* **94**, 14429–14432 (1997).
19. Schlunegger, M., Bennett, M. & Eisenberg, D. Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly. *Adv. Protein Chem.* **50**, 61–122 (1997).
20. Lazaridis, T. & Karplus, M. Multiple unfolding simulations reconcile the “new view” of protein folding with the old. *Science* **278**, 1928–1931 (1997).
21. Roder, H. & Colon, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28 (1997).
22. Frauenfelder, H. & McMahon, B. Dynamics and function of proteins: The search for general concepts. *Proc. Natl Acad. Sci. USA* **95**, 4795–4797 (1998).
23. Burton, R. E., Myers, J. K. & Oas, T. G. Protein folding dynamics—quantitative comparison between theory and experiment. *Biochemistry* **37**, 5337–5343 (1998).
24. Cavagnero, S., Dyson, H. J. & Wright, P. E. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **285**, 269–282 (1999).
25. Ferguson, N., Capaldi, A. P., James, R., Kleinhous, C. & Radford, S. E. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**, 1597–1608 (1999).
26. Taketomi, H., Ueda, Y. & Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulations. *Int. J. Pept. Protein Res.* **7**, 445–459 (1975).
27. Ferrenberg, A. M. & Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195–1197 (1989).
28. Zhou, Y., Karplus, M., Wichert, J. M. & Hall, C. K. Equilibrium thermodynamics of homopolymers and clusters: Molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* **107**, 10691–10708 (1997).
29. Ballew, R. M., Sabelko, J. & Gruebele, M. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA* **93**, 5759–5764 (1996).
30. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of β -hairpin formation. *Nature* **390**, 196–199 (1997).

Acknowledgements

We thank Y. Duan and P. A. Kollman for sending us the structure of the intermediate observed in their trajectory. This work was supported in part by a grant from the NSF and a grant from Pittsburgh Supercomputing Center, and from UC Berkeley Network of Workstations (NOW) through NAPCI (National Partnership for Advanced Computational Infrastructure). Y.Z. is an NIH postdoctoral fellow. The calculations at Harvard were conducted on HP 9000/735, DEC Alpha and SUN UltraSparc workstations. The drawing of the three-helix bundle were made with QUANTA (Molecular Simulations Inc.).

Correspondence and requests for materials should be addressed to M.K. (e-mail: marci@tammy.harvard.edu).

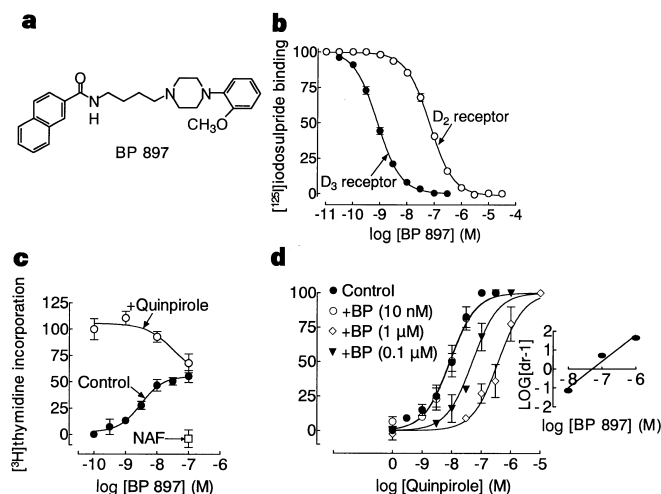
errata

Selective inhibition of cocaine-seeking behaviour by a partial dopamine D₃ receptor agonist

Maria Pilla, Sylvie Perachon, François Sautel, Fabrice Garrido, André Mann, Camille G. Wermuth, Jean-Charles Schwartz, Barry J. Everitt & Pierre Sokoloff

Nature **400**, 371–375 (1999)

Spurious symbols were introduced into Fig. 1d during reproduction. The correct figure is shown here.

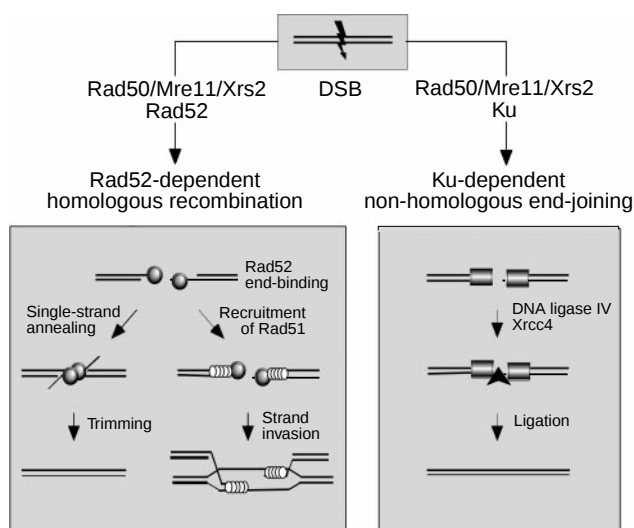


Binding of double-strand breaks in DNA by human Rad52 protein

Eric Van Dyck, Alicja Z. Stasiak, Andrzej Stasiak & Stephen C. West

Nature **398**, 728–731 (1999)

In Fig. 4, one of the strands of duplex DNA was deleted. The corrected figure is shown below. □



E-commerce is a gas

Meet Noggin, and get some wheels for your glassware.

Cuvette Holder

OceanOptics www.oceanoptics.com

Hold on to your cuvettes

The CUV-FL-DA Cuvette Holder attaches directly to the light source and couples via optical fibres to OceanOptics spectrometers to create a small-footprint spectrophotometric system for fluorescence and absorbance experiments. The holder is optimized for UV-VIS-NIR applications, holding 1-cm square cuvettes. Especially useful for fluorescence, the CUV-FL-DA has mirrored screw plugs to make the fluorescence signal stronger. These reflection mirrors are UV-enhanced and aluminium-coated to aid signal reflection. One mirror collects the fluorescence that would otherwise be lost, the other reflects the excitation energy back through the sample.

Reader Service No. 100

WebStore

Adept Scientific www.adeptsience.co.uk

Credit card orders via the web for software

The Adept Scientific WebStore, accessed via <http://directory.adeptsience.co.uk>, makes a wide range of software packages for scientists and engineers available via the now familiar 'shopping cart' system. Brands available include Mathcad, Computer Boards data acquisition products, and the Maple V computer algebra system. The WebStore is closely linked to Adept Scientific's main web site.

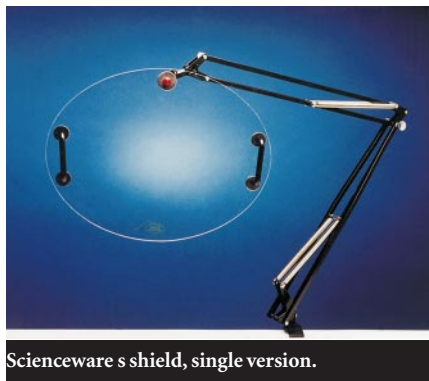
Reader Service No. 101

Splash shields

BelArt Products www.bel-art.com

Adjustable defence

Scienceware Plexiglas Splash Shields are designed for maximum protection between



Scienceware splash shield, single version.

the source of danger and the operator. To free up work space, these units mount on the bench top or against the wall. The shields are available in a single or double design, with side-to-side as well as vertical movement. The Single Splash Shield has a 12 × 15-inch (305 × 381 mm) Plexiglas shield while the Double Splash Shield is 15 × 24 inches (381 × 610 mm). Both units are available with a choice of either a 24-inch (610-mm) or 38-inch (965-mm) arm.

Reader Service No. 102

Gases on-line

Matheson's www.mathesongas.com

E-commerce for speciality gases

Matheson's catalogue of gases and gas-equipment products can now be browsed and purchased online at Matheson's Direct Connection websites. Product literature and application notes are also posted on the site, with most available for viewing and downloading. The Matheson Direct Connection website resides at www.mathesongas.com/catalog. It is a full-featured e-commerce site with product search capability, an editable shopping cart, and fully secured electronic transactions.

Orders can be placed online using a credit card or purchase order number. MSDS sheets are available on-line in HTML format.

Reader Service No. 103

Agrometeorological station

LI-COR <http://env.licor.com>

Transportable environmental monitoring

The LI-1401 agrometeorological station contains a range of LI-COR instrumentation to measure solar radiation, soil and air temperature, relative humidity, wind speed, wind direction and rainfall. It incorporates the LI-1400 DataLogger with an integrated keyboard and display. Input signals can be integrated, averaged and recorded over periods from one second to one day, or a point value can be collected at the end of each period. An RS-232 output and Windows interface software simplify configuration and data transfer for analysis in spreadsheet programs.

Reader Service No. 104

Portable pH meters

Beckman Coulter www.beckmancoulter.com

Hand-held, battery-operated pH meters

The Φ200 Series of three waterproof and dust-tight pH meters could be what you are looking for. All three are hand-held, portable, battery-operated systems capable of multi-point (up to five) calibration. The Φ255 model measures pH as well as mV through oxidation-reduction potential, relative mV and temperature. The Φ265 is a GLP/GMP-compliant system that includes an RS-232 interface to connect to a printer or computer. The Φ295 model offers the capabilities of the Φ265 plus the ability to measure concentration by ion-selective electrodes.

Reader Service No. 105



Measuring pH in the rain?

ADVERTISEMENTS

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes
0.45cm!

READER ENQUIRY NO. 54

Check out our homepage at <http://www.resgen.com>

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 6



Decorative guidance on handling radioactivity.

Safety and radiation

Amersham Pharmacia Biotech
www.apbiotech.com

Being safe with radioactivity

Wallcharts and cards giving guidance on how to work safely with radioactivity and isotopes are now available from Amersham Pharmacia Biotech. These are the first in a series of publications and videos aimed at anyone using radioactivity in the lab. The materials have also been designed to be used as training aids for radiation safety officers. The wallchart explaining the 'ten golden rules' of working with radioactivity is available from local suppliers.

Reader Service No. 106

Glassware Cart

Labconco Corporation www.labconco.com

Move glassware from A to B

The Labconco Glassware Cart is designed to transport a wide variety of glassware, plasticware and small instruments. The welded frame of 1-inch square tubing is coated with corrosion- and abrasion-resistant baked-on epoxy powder. Four-inch-diameter hard rubber casters absorb vibration for smooth, quiet trips over rough, uneven floors. The 19-inch-wide frame fits through most doors and narrow corridors. Glassware Carts are available with two large or four small vinyl-coated wire baskets. A removable plastic drip pan contains spills. The Glassware Cart is delivered ready to roll; no assembly is required.

Reader Service No. 107

Radioactivity shield

Research Products International
www.rpicorp.com

For work with phosphorus-32

This new bench-top beta shield features a contoured profile that allows unrestricted access to the area in front of the shield. Ergonomic design enhances dexterity and reduces fatigue when using the shield for extended time periods. The shield is made of 3/8-inch-thick acrylic for shielding high-energy beta emitters such as phosphorus-32. Overall shield dimensions are 15 inches wide × 21 3/4 inches high.

Reader Service No. 108

Nuclear waste furnace

Carbolite www.carbolite.co.uk

Quantitative waste handling

A new furnace developed by Carbolite and AEA Technology provides convenient and reliable combustion of nuclear waste to create samples for carbon-14 and tritium analysis. The equipment combusts the sample material to completion and traps the principal products — carbon dioxide and water — for assessment by liquid scintillation counting. The samples are held in a removable combustion tube which can be replaced quickly and easily.

Reader Service No. 109

Lab Notebooks

Nalgen www.nalgenunc.com

Get the evidence in writing

These new Laboratory Notebooks contain 200 pages of acid-free paper pages with 1/4-inch grid lines. The 8 1/2 × 11-inch pages are permanently bound and sewn. Page format includes places for preparer and witness signatures to aid in securing patent protection. The cover is in a spill-resistant material. Acid-free recycled paper pages resist yellowing and ageing. Preprinted instructions and documentation signature requirements conform with Good Laboratory Documentation.

Reader Service No. 110



Noggin GPR, ready to roll.

Ground-breaking radar

Sensors&Software www.sensoft.on.ca

'Intelligent' subsurface imaging

Latest refinement for the Noggin ground-penetrating radar (GPR) system is the Noggin Smart Cart. This exploits 'next generation' technology to enable users to locate areas of subsurface interest without a-priori information. The cart incorporates a digital video logger which immediately displays subsurface images on screen as data are being collected. The Smart Cart is also available as an accessory for Noggin 250 and Noggin 500 GPR system owners.

Reader Service No. 111

BAS-1800II

Fujifilm www.fujimed.com

An addition to the company's Bio-imaging Analysis System

Designed to cope with high-throughput macroarray image analysis, the BAS-1800II is small but has the power necessary for a wide range of procedures. The unit is recommended for Southern, northern and western blotting, *in situ* hybridization, electrophoresis, X-ray diffraction and other techniques. Macroarray analysis software (ArrayGauge) is available for use with the various Windows operating systems.

Reader Service 112

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

**HIGH QUALITY
PEPTIDE SYNTHESIS**

- 'Standard' and 'non-standard'
- Fast delivery
- Competitive prices
- Also advisory services

Pepsyn Ltd

Contact: Dr J A Smith, Chairman, Pepsyn Ltd,
 School of Biological Sciences, Life Science Building,
 The University of Liverpool L69 7ZB, UK
 Tel: 0151 794 4329 Fax: 0151 794 4349
 e-mail: J.A.Smith@liverpool.ac.uk
 Website: www.pepsyn.co.uk

READER ENQUIRY NO. 22

**Detect residual DNA with Wako's
DNA Extractor Kit
or Isolate Genomic DNA with
DNA Extractor WB Kit**
 Both use the Nal Method
 No Phenol No Chloroform

Wako

Wako BioProducts
 Wako Chemicals USA, Inc.
 1600 Bellwood Road
 Richmond, VA 23237
 Telephone: (804) 271-7677
 Facsimile: (804) 271-7791
 (800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 39

NEWCASTLE UNIVERSITY
 MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING
Membrane-bound, solid or solution samples
£12 per cycle
 5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray:
 Phone/Fax: +44 191 222 8612
 Email: JOE.GRAY@newcastle.ac.uk

*DNA sequencing (ABI 377-96KL, 877 Catalyst)
 also available at competitive prices*

READER ENQUIRY NO. 2

CLIMBING THE LADDER

Four of science's leaders — the head of a university college, and the directors of a medical research

charity, an observatory, and a pharmaceuticals company — tell how they made it to the top
pp408–409

FROM DESK TO BENCH

A few dynamic administrators find the time to keep their hand in at scientific research — and

discover that lab work gives them added credibility when it comes to decision-making
p410

How head-hunters track down the winners for science's top jobs

The star performers are not just head-hunted, they can write their own job descriptions. But how can a scientist reach this enviable position?

Business is booming for head-hunters. More and more directors of industrial research and development, heads of research organizations and university vice-chancellors are being selected through executive search agencies — companies that typically charge £35,000 (US\$56,200) to find someone capable of filling a top job with a salary of £100,000.

According to Kennedy Information — which calls itself “the leading information source on management consulting and executive search agencies” — the worldwide revenue of the executive search industry has grown rapidly, from around \$3.5 billion in 1993 to \$7.3 billion in 1997.

One of the biggest players is Korn/Ferry International, specializing in areas including advanced technology, education, energy and health care. It has 70 offices in 40 countries, employing about 1,500 staff who serve more than 4,000 clients. It floated on the New York stock exchange this year. Half its revenue comes from the United States, 30 per cent from Europe, 10 per cent from Asia and 10 per cent from Latin America.

Defining the job

The first task of an executive search agency is to identify exactly what a job entails. “For every appointment, we have a briefing meeting with the client to learn about the job and the working environment, and to identify what a successful appointment would look like,” says Peter Bassett, UK managing partner of health-care practice at Korn/Ferry International in London. “We ask what success would look like at six months and at three to four years. By asking the client the question in that way it helps them define their thinking and therefore the profile of the ideal candidate. We ask why candidates would want to do the job and what the challenges are.”

Having identified the job description, “The next question is: where would the person who could do that job be now?” says Bassett. “They could be in academia or industry. We can consult a database and conduct a

search by contacting people who have helicopter vision on who is new and upcoming. Another way is to look at an organization and identify who does what within it.”

Most executive search agencies rely on sources to suggest names. Many senior researchers are accustomed to being contacted by agency consultants, most of whom have spent around ten years working in the com-



Llewellyn Smith:
“agencies often call”.

Agencies typically contact up to 150 people, gathering informal references and getting a feel for the remuneration that would be required. “We confirm a recommendation by careful reference taking,” says Bassett. The agency passes a shortlist of around 12 names to the client for interview.

Executive search agencies first appeared in the United States about 50 years ago. Egon Zehnder has been around for 35 years, and has 50 offices in 40 countries. It specializes in areas including the life sciences, chemicals and high technology.

“There are far more similarities than differences between the use of executive search agencies in the US and Europe and Asia,” says Ron Tracy, a consultant in the firm’s life sciences practice group covering the United States, a larger and more heterogeneous market than the United Kingdom. In Asia, he says, many companies remain local, but there are an increasing number able to compete in regional and international markets, so recruitment will become more international.

As recruitment searches become more

international, the ability to adapt to different cultures is becoming vital to success in a senior position, according to Laurence Monnery, a consultant in Egon Zehnder’s European life sciences practice group. “Searches are international as the candidate could be anywhere,” she says. “The top researchers are not necessarily the best managers and leaders. Whatever the technical skills required, we will always look at other aspects — people management, leadership and the ability to adapt to different cultures.”

A place on the board

Egon Zehnder also recruits non-executive board members for biotech companies. “Companies are placing more emphasis on having strong independent board members, and we look to recruit these from academia. It is important that the candidate has the ability to stand back, advise and challenge.”

Founded in 1953, Heidrick and Struggles was one of the first executive search agencies. It now has 59 offices around the globe, with clients including start-up companies, universities and hospitals, and a revenue of \$330 million (including subsidiary firms) in 1998. Although most executive search agencies do not provide recruitment services for individuals seeking work, Heidrick and Struggles has recently started an individual placement service for technology professionals.

An alternative for researchers seeking a new lease of life is a temporary placement as a troubleshooter. “There are a variety of scenarios in which we are called in,” says Cynthia Larbey, managing director of People in Health, a British firm based in Woodbridge,

Useful websites

- Korn/Ferry International <http://www.kornferry.com>
- Egon Zehnder International <http://www.zehnder.com>
- Heidrick and Struggles <http://www.heidrick.com>
- Saxton Bampfylde <http://www.hevergroup.com>
- Association of Executive Search Consultants <http://www.aesc.org>

Suffolk, which operates across Europe. "For example, we had the medical director of a biotech company which was coming up for a product launch off with a long-term illness at the most crucial time for the company. Fortunately we had a chap with experience in the right area on our books." People in Health has more than 130 people poised to take a temporary job. "They are people who have taken early retirement or been made redundant but who aren't ready to take up the golf clubs," says Larbey. "Plus 40- to 45-year-olds who have made a lifestyle choice over the way in which they want to work. But I don't employ people between jobs. Only five per cent of those who approach the company are accepted."

Although industry is still more likely to use head-hunters than academic institutions, the public sector does use them to fill senior positions. "We have placed some 30 university vice-chancellors in the UK in the past three or four years," says Paula Alexander of Saxton Bampfylde. "The qualities which clients seek vary hugely. Some universities are looking for a real research person, as their research rating is poor and they want to get it rocketing. Others are looking for a business person as they have financial problems."

The Wellcome Trust, the world's largest medical research charity, uses executive search agencies. "Most organizations are moving towards using head-hunters," says Mike Dexter, the trust's director. "Head-hunters can make phone calls which the appointing organization can't without it appearing to be a firm job offer. Head-hunting is an informal peer-review process but it is an informed peer-review process. There is a selection informed by debate."

Of course there are ways to secure a senior appointment without using head-hunters. Organizations can appoint their own search committees that can use the same techniques as head-hunters. And in many cases the ideal candidate is obvious and can be approached direct. Some companies identify who they want and design a job in collaboration with the person they recruit. Peter Goodfellow, who runs SmithKline Beecham's drug discovery programme, secured his job in this way.

Chris Llewellyn Smith

University College London

Chris Llewellyn Smith became provost and president of University College London (UCL) five months ago. Before that, he was director-general of CERN, the European Laboratory for Particle Physics, in Geneva.

Hailed as an outstanding diplomat, Llewellyn Smith had been credited with laying the foundations for European particle physics by persuading governments to fund the Large Hadron Collider, due for completion in 2005. He convinced countries that were not members of CERN, in particular

the United States and Japan, to help fund its construction. Llewellyn Smith also oversaw a scheme to cut pay by allowing staff to exchange a percentage of their salary for extra leave, freeing up enough money to recruit new young staff.

Llewellyn Smith knew that Derek Roberts, the former head of UCL, was about to retire. He, too, was due to leave his job at the end of his five-year term. As director-general of CERN, he had overseen an organization with an annual budget of £375 million and a staff of 2,800 plus around 6,500 scientific users from universities and research institutions in 50 countries.

After the universities of Oxford and Cambridge, UCL receives the biggest grant for research from the Higher Education Funding Council for England. This money, more than £60 million a year, helps to lever income from the research councils, industry and charities.

"I was talking to colleagues based at UCL and they suggested that I might be interested in becoming provost," says Llewellyn Smith. "I then spoke to my predecessor, Derek Roberts, to decide whether or not I was interested in the position." UCL employed Korn/Ferry International to discover who might be interested in the job. A month after his conversation with Roberts, the head-hunters approached Llewellyn Smith and told him that his name had come up during talks to identify suitable candidates. The agency then sent him more information about the position and encouraged him to apply.

Llewellyn Smith had three lunches with eight people to discuss the job, his experience and his views on the role of provost. This was followed by a day of formal interviews, after which he was offered the job. His salary of about £135,000 represents part of a package with a net value comparable to what he was getting at CERN. Being provost of UCL is a distinctly different job, Llewellyn Smith finds. "CERN is project-based and the scientific policy had to be led by the director-general," he says. "Within a university, academic policy is set from the bottom up and my job is to create the right conditions for research — encouraging rather than directing."

Llewellyn Smith now spends much time dealing with funding proposals from the UK government agencies that allocate money to universities. "The funding council breaks up funding into chunks which it gives for schemes such as collaborating with industry on research or for widening access to higher education to people from disadvantaged backgrounds. We feel micromanaged from the centre," he complains.

What advice would he give to a scientist aspiring to a senior appointment? "Think carefully," he says. "It is very difficult to carry out work of your own. However there are compensations, such as learning about other areas of science."

Mike Dexter

Wellcome Trust

Mike Dexter became director of the Wellcome Trust, the world's largest medical research charity, in June 1998. Before that, he was head of the Paterson Institute in Manchester, United Kingdom, where he had spent much of his working life.

"I did know that the position was coming up because I had been approached by a head-hunter who wanted me to suggest names," he says. "I told him I wasn't interested as I had recently accepted the job as director of the Paterson Institute." The Wellcome Trust then approached Dexter to seek his views on the future directions of biomedical science. "At that stage, I had no idea that they had a view to my becoming director," he says. "I was very relaxed because it was not a job interview. We discussed the Wellcome Trust's activities and



Dexter: 'once-in-a-lifetime opportunity'.

positioning in UK science, career development, clinical research, university infrastructure and other research institutes. I suppose I was interviewed without knowing it. After our conversation, the governors discussed what they were looking for in a director and approached me again." Dexter had been based at the Paterson Institute almost continuously for 35 years. He started in 1963 as a junior research technician, taking a three-year break to get his degree as a mature student, as illness had prevented him from entering university straight from school. He then completed his PhD at the institute and worked his way up. "Every time I thought about moving, the Paterson Institute came up with a better offer," he says.

However, the Wellcome offer was one that the Paterson could not better. "It was an opportunity that comes up once in lifetime and it was the only job that I would leave Manchester for," says Dexter. "The Wellcome Trust is an organization which has an impact not only on British biomedical science but internationally. As director I could have a strategic input into where biomedical science is going."

Two weeks after Dexter succeeded Bridget Ogilvie as director, Wellcome ploughed £300 million into refurbishing decaying university labs and £100 million towards a new synchrotron. The UK government simultaneously pledged an extra £300 million for this refurbishment and £300 million for the synchrotron. Further government money, directed through the research councils and the Department for Education and Employment, took the total increase to £1.4 billion over three years. Dexter and his predecessor

have been credited with leveraging this public money using Wellcome Trust funds.

"It was an exciting time," he says. "I thought it couldn't always be like that but I have been involved in ongoing initiatives. The job is so varied. One day I can be meeting senior politicians at ministerial level, on others the chief scientific adviser to the government, the director-general of the research councils, the heads of the research councils, and overseas scientists."

Money is not a prime motivator, says Dexter. Indeed, other scientists believe that his salary of between £130,000 and £140,000 is modest for what Dexter could command. But Dexter has boosted the pay of his researchers by up to 30 per cent, and has called on the government to increase pay for university researchers. "We can only exploit people so much," he says.

The only downside to the job is that Dexter can no longer run an active research group. "The hardest decision was to move over into what is essentially administration," he says. "I certainly would not encourage a young researcher to go into administration while they are still doing their best research. I miss doing science but it is enormously exciting to develop strategies and policies for biomedical science internationally."

Catherine Cesarsky

European Southern Observatory

Catherine Cesarsky became director-general of the European Southern Observatory (ESO), in Garching, near Munich, this month. Before that, she was head of the department of "science of matter" at the French Atomic Energy Commission (CEA), conducting basic research in physics, chemistry, astrophysics and Earth sciences.

"I had a phone call from the ESO search committee, which had been appointed to find a new director-general to succeed Riccardo Giacconi," says Cesarsky. The committee had drawn up a shortlist of suitable candidates after consulting key figures in the astronomy community.



Cesarsky: 'this is an exciting adventure.'

Cesarsky's high profile had ensured her presence on the shortlist — she is vice-president of the International Astronomical Union, vice-president of the European Space Agency council and principal investigator for one of the experiments on board the agency's Infrared Space Observatory, and editor-in-chief of *Astronomy and Astrophysics*. In fact, her profile is such that a full-page article about her appeared in last December's French edition

of *Vogue*, dedicated to the theme of science and fashion.

Each of the eight member states of ESO sent two representatives to interview Cesarsky simultaneously. "There was no-one on the board that I had not met before, they were all people with whom I have had working interfaces, so it was not intimidating at all," she says. After an interview lasting one hour, she was offered the job. One of the conditions set by Cesarsky was that she should still be able to conduct research. "I have to have my science to be happy," she says. "If I am happy, then I can manage people well."

About 400 people work for the ESO, which has a budget of DM1.5 million (US\$800,000). When Cesarsky worked for the CEA, she was responsible for some 3,000 staff and had a budget of FF2 billion (\$317 million). Yet she is getting paid more at the ESO. "I was well paid at the CEA and I have received a pay rise in coming to ESO," she says. "However it is offset by extra expenses: my children are in France and I need to maintain a house there." She estimates her salary as £90,000.

Her international lifestyle propelled her towards accepting the job. Born in France, she was raised in Argentina and completed her PhD and a postdoc in the United States. However, she had been living in France for the past 25 years and was keen to move on. "My husband and I are excited by change," she says. "Our children are now at the age when they can be left alone and my parents both died recently. This job represents a new lease of life, a new adventure and a new opportunity. I have a special fondness for ESO. It is a wonderful time, to see what will come out of the Very Large Telescope."

Peter Goodfellow

SmithKline Beecham

Peter Goodfellow joined SmithKline Beecham three years ago and now runs its drug discovery programme. Before that, he was professor of genetics at the University of Cambridge.

SmithKline Beecham created a job specifically for Goodfellow, a renowned expert in human genetics, to tempt him to join the company. "I was approached by George Post, who was then head of research and development at SmithKline Beecham," says Goodfellow. "We met at a conference. I was walking down a corridor and he asked me whether I was interested in joining SB. I visited the company sites in the US and UK, and we worked together to define a job that was appropriate."

Goodfellow had gained his first degree at the University of Bristol, in the same department in which Post had completed his PhD. "There was a cultural history," says Goodfellow. "Post knew who I was and I knew who he

was." After graduating, Goodfellow went to the University of Oxford to do his PhD, followed by three years at Stanford University, California, and 13 years at the Imperial Cancer Research Fund in the United Kingdom.

During his time in the United States, he gained some industrial experience and, after returning to the United Kingdom, he founded Hexagen as a spin-off company from the University of Cambridge. In August 1998, Hexagen was bought by the US company Incyte Pharmaceuticals for £23.5 million. Having a 2 per cent stake in Hexagen, Goodfellow made the best part of £500,000.



Goodfellow: 'I feel like a student again.'

SmithKline Beecham were interested in my skill base in human genetics but I was keen not to be its 'expert in genetics' — I wanted to make drugs," says Goodfellow. He was taken on to run a platform of genetics technologies and within three years was promoted to head of research. Responsible for some 2,000 scientists on three sites, he is cagey about his earnings. People working in similar jobs are likely to earn up to £175,000 a year.

"The way in which knowledge is transformed into doing good for people is my major motivation in working for the company," he says. "There is something of a buzz about working in health and welfare." Goodfellow views his responsibility for areas of science outside his speciality as a bonus. "It is wonderful to have the opportunity to learn across such a wide area of expertise. Most of the things I have responsibility for I know little about. I feel like a graduate student again."

About 30 per cent of Goodfellow's time is spent on strategic issues within research and development and the company as a whole. A further 30 per cent, he estimates, is spent reviewing science, and 30 per cent administering his group. Goodfellow devotes his remaining time to academic research and other responsibilities — he is a trustee of the Imperial Cancer Research Fund, for example. "You have to make a decision about how close you want to be to the everyday science," he says. "In a senior position, it is not possible to spend the same amount of time doing research. There are people who can still run a lab but it can't be done as if it were the only priority. You have to accept the fact that you are going to spend time dealing with the administration and politics of an organization — there are a lot of interfaces at which you need to work."

Alison Goddard

Alison Goddard is a reporter on The Times Higher Education Supplement.

e-mail: alison.goddard@newsint.co.uk

Keeping in touch with research

Moving into a position of leadership or administration does not necessarily mean moving away from research. More scientific leaders are following the precedent set by Harold Varmus, David Baltimore and Daniel Koshland, and maintaining active roles in their research while heading scientific organizations.

This is the case for Nobel laureate Thomas Cech, due to take over as president of the Howard Hughes Medical Institute (HHMI) from Purnell Chopin at the end of the year, as well as continuing to lead his research group at the University of Colorado at Boulder. HHMI initially approached Cech to solicit his help in its search for a new president and subsequently asked whether he would consider the position himself.

Whether by design or not, this turned out to be a good approach to recruiting Cech, who in his 25 years on the faculty at Colorado declined to consider offers of university presidencies and other leading roles at research



Cech: will commute between two jobs.

institutions, turning away the many head-hunters who came knocking at his door. Cech's intention was to wait until his children finished high school. Other factors, such as the political nature of the work and the large amount of fundraising expected of a president, made such jobs seem less than attractive. But the HHMI presidency is a unique opportunity, says Cech. "The current president attends all the scientific meetings, as well as the renewal reviews of the investigators, and there is no fundraising involved — Howard Hughes did that for us."

During his interviews with HHMI, Cech told the search committee that his interest in the job would only hold if he could maintain a research programme. So, like other organizations that recruit top scientists from their research labs, HHMI has written into Cech's employment agreement that he is expected to maintain his Boulder laboratory. "I thank my friends Harold Varmus, David Baltimore, Dan Koshland and others for having blazed this trail," says Cech, who will be commuting each month between Colorado and HHMI's headquarters in Chevy Chase, Maryland.

No stranger to wearing a variety of hats, Rita Colwell is used to balancing research with other roles, first as an administrator at the University of Maryland, then as chief academic officer there, and now in her second year as the head of the US National Science Foundation (NSF). "It's not like you simply

leave academia," she says. If anything, she believes her post at the NSF has broadened her contacts with academic research. She now works with astronomers, geoscientists and engineers, but in continuing her research on cholera in her own time she can



Colwell: 'research helps in leadership'.

stay in contact with her students and with scientists in her own field. "It's always difficult when you have something you love so much that you want to put 100 per cent of your time into it, and 200 per cent of your time is a bit difficult," she says wryly. However, she contends that this approach adds authenticity to her leadership: "There is a great deal of credibility that goes with your decisions, especially when you are still in a day-to-day [research] operation."

Cech sees it from a slightly different angle: "In a leadership position, one has to find some way to keep in touch with what is happening at the base of the pyramid. Remaining active in research is not the only way to do this, but it seems like it could be a very successful way." His decision seems to have been well received by his peers. He says there has been wide support for his decision to maintain a research programme. "There may be some people who think it's a conflict of interest or an unrealistic goal, or are upset about it. If so, they have been politely silent," he says.

According to Mary Clutter, an assistant director at the NSF, there are other advantages to taking on a leadership role in science, especially earlier on in one's career. Clutter directs a group of programme officers in the directorate of biological sciences and says that "people don't realize why these jobs might be helpful to their own careers". Even mid-career researchers who have established themselves in a discipline do not know how useful it is to go to Washington DC, to be in charge of a whole area and to see every proposal that comes in for review, she says. "Your whole outlook on life broadens. You understand from the larger perspective your own research area. You learn about techniques that you might use in your own research that you had never thought of because they come from another discipline. It's unbelievably valuable."

For the aspiring leader, becoming involved with foundations, societies and other organizations can increase one's exposure and result in mentoring opportunities, which can contribute significantly to a person's upward mobility. According to Clutter, mentoring did not develop as a formal con-

cept until about 20 years ago, and it may have arisen because women complained about unequal opportunities. "Twenty years ago, mentoring was spontaneous — people didn't even know they were doing it, and it was much more common among men than among women," she says. Today, if there are only a few women in an organization then there is less mentoring among women, says Clutter. "Women may feel they are competing for 'the one' position that may be available for women." The same principle may hold for minority groups, she adds. It is Clutter's policy to assign mentors to all those in her directorate, from support staff to the most senior people. She is convinced that it is an effective strategy, providing support and professional friendship for those in training.

Mentors open doors

"The two best mentors that I've had since I came to the federal government were men," says Clutter. The first, who was from another agency, introduced himself to Clutter when she was a programme officer at NSF. Later, when she was in a more senior position, it was a director of NSF who acted as a mentor, although he wouldn't have thought of it that way, she says. "He introduced me to everybody — it was amazing how much I learned." Clutter credits that relationship with opening the doors that led to the assistant directorship of NSF. As she puts it, she already had the necessary qualities and qualifications, but the mentor provided the introduction to the people who needed to know what she could do.

The idea that we stay in one career or one job for our whole working life has given way



Clutter: 'effective mentoring pays off'.

to the idea that we can change careers several times before we retire, says Colwell. "I think that's the most dramatic challenge." Mentoring is an important part of this process. For Colwell's students, particularly the women, she is living proof that there are no limits to what someone can achieve.

And even though the exceptionally busy life she leads may make her less accessible than many advisers, Colwell believes that the time she spends with her students is quality time. "I am able to draw on a lot of experience and expertise to ensure that the path that they take in their research is productive and will pay off in findings that are of interest and timely."

Brendan Horton

*Brendan Horton is a freelance science writer.
e-mail: brenhorton@aol.com*